AD-A105 090 HASKINS LABS INC NEW HAVEN CT F/G 17/2

STATUS REPORT ON SPEECH RESEARCH, 1 APRIL-30 JUNE 1981.(U)

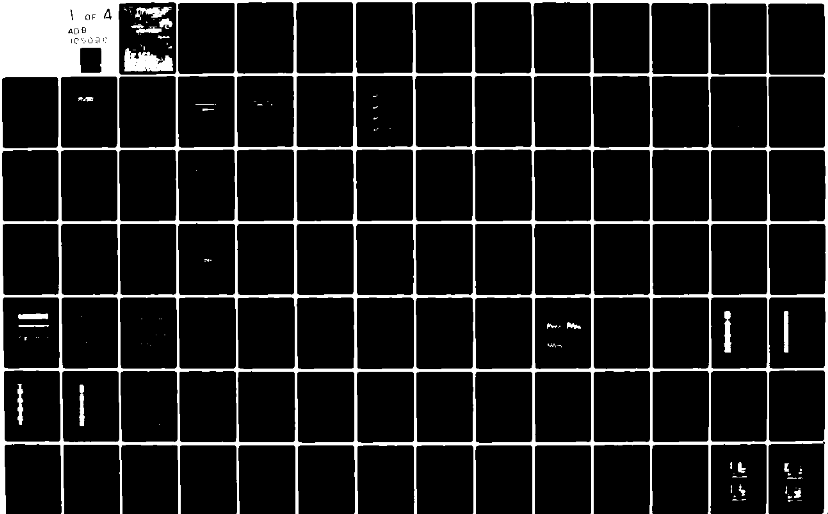
1981 PHS-N01-HD-1-2420

UNCLASSIFIED SR-66(1981) NL

1 OF 4

ADB

105090



SR-66 (1981)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 April - 30 June 1981

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

National Institute of Child Health and Human Development
Contract N01-HD-1-2420 ✓

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant MCS79-16177
Grant PRF8006144

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617

National Institute of Education
Grant G-80-0178

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter Alfonso*
Thomas Baer
Fredericka Bell-Berti*
Catherine Best*
Gloria J. Borden*
Susan Brady*
Robert Crowder*
Carol A. Fowler*
Louis Goldstein*
Vicki Hanson
Katherine S. Harris*
Alice Healy*
Kiyoshi Honda[1]
Daisy Hung
Leonard Katz*
Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Charles Marshall
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Ovid Tzeng[2]
Mario Vayra[3]
Robert Verbrugge*

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Vincent Gulisano
Donald Hailey
Terry Halwes
Sabina D. Koroluk
Agnes M. McKeon
Nancy O'Brien
Marilyn K. Parnell
Susan Ross*
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

Claudia Carello
Tova Clayman
David Dechovitz
Steven Eady
Jo Estill
Laurie B. Feldman
Carole E. Gelfer
David Goodman
Janette Henderson
Charles Hoequist
Robert Katz
Aleksandar Kostić
Peter Kugler
Anthony Levas
Harriet Magen
Suzi Pollock
Patti Jo Price
Sandra Prindle
Brad Rakerd
Daniel Recasens
Rosemarie Rotunno
Arnold Shapiro
Susan Smith
Rosemary Szczesiul
Louis G. Tassinary
Douglas Whalen
Deborah Wilkenfeld

---

*Part-time
[1]Visiting from University of Tokyo, Japan
[2]Visiting from University of California, Riverside
[3]Visiting from Scuola Normale Superiore, Pisa, Italy

SR-66 (1981)
(April-June)

CONTENTS

I. Manuscripts and Extended Reports

I. MANUSCRIPTS AND EXTENDED REPORTS

ELECTROMYOGRAPHY AS A TECHNIQUE FOR LARYNGEAL INVESTIGATION*

Katherine S. Harris+

While, as earlier papers at this conference have indicated, the forces that determine laryngeal adjustment are complex, muscular forces are extremely important. In recent years, techniques for studying muscle activity in general have improved, and with these developments, the study of the laryngeal muscles in normal alert humans has become possible using the techniques of electromyography. In this paper, I will discuss some properties of muscles, and of the laryngeal muscles in particular, techniques for EMG recording, and, finally some results of studies on the muscular control of the larynx.

MUSCLE PROPERTIES

The building block for a consideration of muscle activity is the motor unit. This term was coined by Liddell and Sherrington (1925) to include the motoneuron and the muscle fibers it supplies. The contractile response to one impulse in one motor neuron is a twitch contraction in the innervated muscle fibers. Thus, the smallest unit of muscular activity is a contraction of the muscle fibers of a single motor unit, and the smoothly graded contraction of a muscle is accomplished by temporal and spatial summation of the activity of a number of motor units.

The muscles of the body have somewhat different tasks, and their properties are well-correlated with these tasks. For example, some muscles, such as the muscles of the fingers, must make finely tuned movements, while others, such as those of the leg, must support the body against the forces of gravity for long periods of time. These muscles differ in the size of their motor units, and in the histochemical properties of the individual muscle fiber properties that determine their resistance to fatigue.

Table 1 presents some data on motor unit size in the intrinsic laryngeal muscles, with data on one of the eye muscles and the biceps for comparison.

---

*A version of this paper was presented at the Conference on Asssessment of Vocal Pathology, Bethesda, Md., April 1979. (Proceedings to be published in ASHA Reports.)

+Also Graduate Center, City University of New York.

Acknowledgment. This work was supported by NINCDS Grants NS13870 and NS13617, and BRSG Grant RR05596.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

Table 1

Data on the Innervation Ratio of the Intrinsic Laryngeal Muscles, with Some Comparison Information on One of the Eye Muscles and the Biceps

Source	Larynx					Other	
	CT	TA	IA	PCA	LCA	Rectus Oculi Lateralis	Biceps
man (Faaborg-Andersen, 1957)	166		247	116			
man (English & Blevens, 1969)	30						
cat (English & Blevens 1969)	55	90		64			
man (Buchtal, 1973)						13	750

CT Cricothyroid
TA Thyroarytenoid
IA Interarytenoid
PCA Posterior cricoarytenoid
LCA Lateral cricoarytenoid

While different authors have found differences in the number of fibers in a motor unit, there is a general agreement that the laryngeal muscles have low innervation ratios, though not quite so low as those of the eyeball and middle ear; the muscles of the limbs and trunk have generally far higher ratios.

The muscle fibers themselves consist of a number of myofibrils, made up, in turn, of a parallel, overlapping array of actin and myosin filaments. In contraction, the actin and myosin filaments slide relative to each other, so that the muscle shortens and develops tension. In normal physiological conditions, this shortening is initiated by the release of a chemical transmitter, acetylcholine, at the nerve-muscle junction, the motor end plate.

When a muscle fiber is at rest, there is a potential difference across the cell members of about -90 mV, due to the difference in its permeability to sodium and potassium ions. When a nerve impulse reaches the motor end plate, acetylcholine is released, which changes the permeability of the membrane to sodium and potassium ions. If this depolarization reaches sufficient levels, the change in potential becomes self-regenerating, and travels along the muscle fiber. During the passage of this action potential, the membrane potential rises, then reverses its sign and finally returns to its resting value of -90 mV. The movement of ions, and the associated changes in potential, are, of course, the events generating the electromyographic signal. The ionic currents at the membrane apparently release calcium ions within the muscles; the diffused calcium activates the contractile component of the muscle, producing the mechanical effect of muscle shortening or tension development (Carlson & Wilkie, 1968).

While the fibers of striated muscles share many properties, they show some adaptations to their individual tasks. The muscles of the larynx must be well designed for rapid adjustment; however, because of their participation in respiration, they must have some capacity for sustained activity without fatigue. Muscle fibers are of two basic types, red and white, although there are variants in different systems in different animals. The "red" and "white" designations refer to a difference in the fiber color, familiar from the light and dark meat of chicken. The two types differ in their metabolic properties, with red muscle more suited to sustained contraction due to the fatigue resistance and white more suited to rapid phasic contraction. Most muscles of the body, including the muscles of the larynx, show mixed red and white fibers. Any single motor unit, however, is composed of fibers of a uniform type (Brandstater & Lambert, 1973) although, since adjacent motor units have overlapping territories, a cross-section of a muscle will show a checkerboard pattern of red and white.

Biochemical and histological studies of the laryngeal muscles to that date (1970) were summarized by Sawashima. He concluded that, with respect to metabolic properties, the intrinsic laryngeal muscles as a group appeared to be intermediate between skeletal and heart muscles. However, he found disagreements among the authors he reviewed as to similarities and dissimilarities within the group.

Since that review, there have been further studies of the histochemistry of the intrinsic muscles of the larynx. Data from one of them (Edström, Lindquist, & Mårtensson, 1974) are shown in Table 2, showing the percentages

Table 2

Data on Histochemical Properties of the Intrinsic Laryngeal Muscles in Cat, after Edström, Lindquist, and Mårtensson (1974)

	TYPE I		TYPE II		
	(1)	(2)	(1)	(2)	(3)
Fiber type in skeletal muscle (Kugelberg, 1973)	-	I	-	IIA	IIB IIC
Overall % in laryngeal muscles, with most common subtype starred					
CT	40%	*		60%	*
TA	* 10%		*	90%	
PCA	* 40%			60%	*
LCA	* 10%		*	90%	

Table 3

Data from Atkinson (1978) on the Mean Response Time for Some Intrinsic and Extrinsic Laryngeal Muscles

	Intrinsic Laryngeal Muscles			Strap Muscles	
	CT	TA	LCA	ST	SH
Mean Response Time	40	15	15	120	70

of Type I and Type II (red and white) fibers found for each of the four laryngeal muscles examined. While some of the fibers were like Type I and Type II fibers found in limb muscles, others were variants of previously identified types. It is interesting to note that Type II variants are far more common in the thyroarytenoid than in the cricothyroid.

A second study (Sahgal & Hast, 1974) examined the histochemical reactions to ATP and three oxidative enzymes in cricothyroid and thyroarytenoid. The results show some differences between the muscles, which the authors believe are also related to the differences in the speed of contraction of the muscles.

Thus, differences in the histochemistry of the muscles appear to be reflected in their contractile properties. We have seen that the laryngeal muscles are composed predominantly of Type II fibers, like the intraocular muscles in man (Kugelberg, 1973). The laryngeal muscles are generally agreed to be fast muscles, although different authors have obtained different values for their contraction time, the time from nerve or muscle stimulation to the peak of the muscle tension. Figure 1, adapted from Sawashima's review (1970), summarizes the results. The thyroarytenoid is consistently found to be faster than the cricothyroid, which is consonant with the difference in proportion of Type II fibers in the two muscles and, according to Sahgal and Hast (1974), with the difference in their histochemical properties.

Contraction time for the intrinsic laryngeal muscles has been estimated by a very different technique by Atkinson (1978) at Haskins Laboratories. He reasoned that if a causal relationship between f_0 and the EMG activity of various laryngeal muscles were assumed, there should be a correlation between f_0 and gross EMG activity, at some time delay determined by the mechanical properties of the muscle. Thus, cross-correlation analysis should provide clues to relative contraction time.

He asked speakers to produce sentences varying in stress and intonation, thus varying f_0, and cross-correlated average f_0 and rectified and averaged EMG activity, at varying delay times. Table 3 shows the delay times at which correlation reached peak value for different muscles. The finding of shorter mean response time for thyroarytenoid and lateral cricoarytenoid than for cricothyroid, with longer response times for the strap muscles, is like the results obtained by more conventional techniques, summarized in Figure 1, and also parallels the histochemical grouping of TA with LCA, shown in Table 2.

THE ELECTROMYOGRAPHIC SIGNAL

The origin of the electromyographic signal is discussed above in only very general terms. If the signals from the laryngeal muscles are to be considered in detail, the recording procedure itself must be discussed. Figure 2 (Geddes, 1972) shows a muscle with a pair of recording electrodes on its surface. The fibers are aligned parallel to each other. When a muscle fiber or the nerve is stimulated, a wave of depolarization passes along each stimulated fiber. However, since each recording electrode is most sensitive to the fiber closest to it, the event recorded will be weighted by the distance between the pickup and the active fiber, as shown in the figure. As

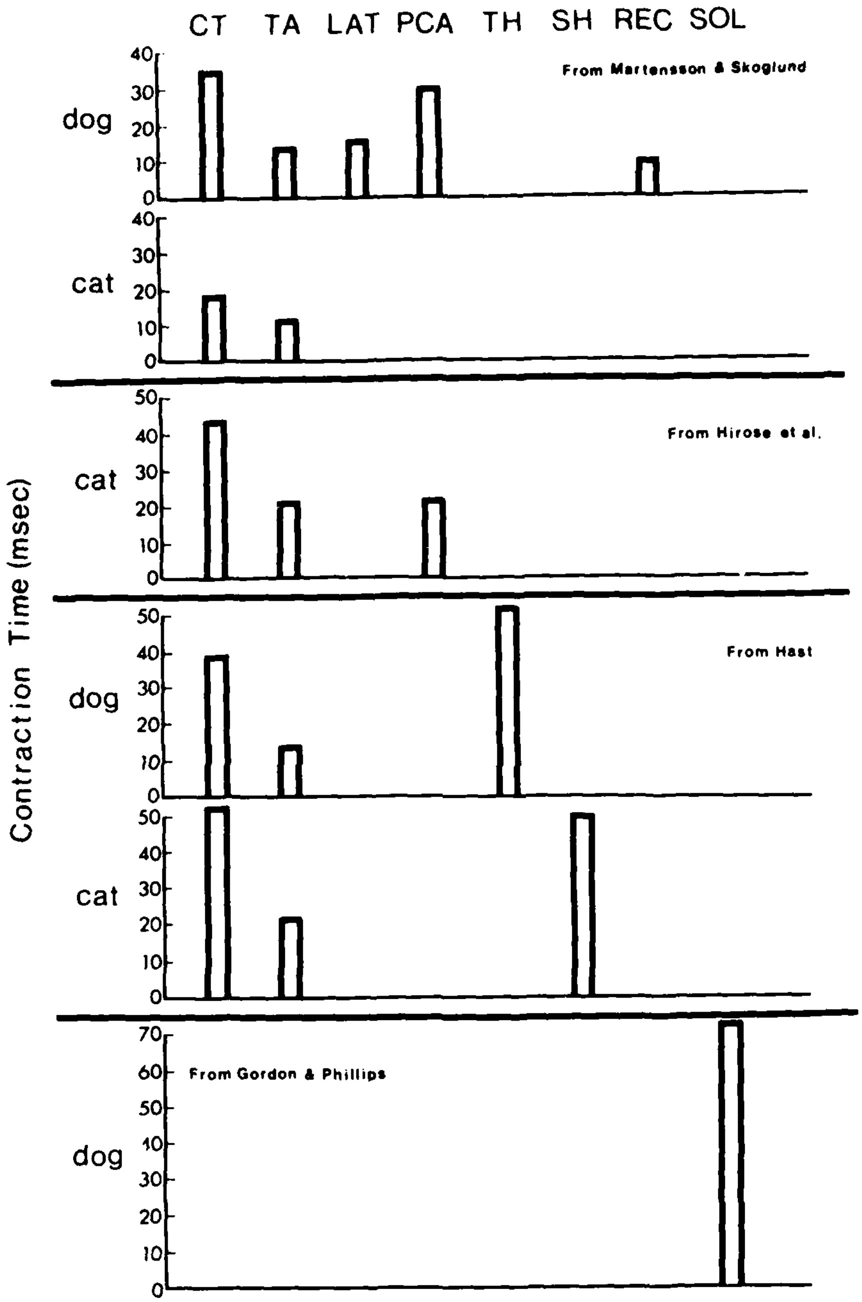


Figure 1. Contraction time in msec for various laryngeal muscles. This figure is adapted in part from Table 1, Sawashima, 1970.

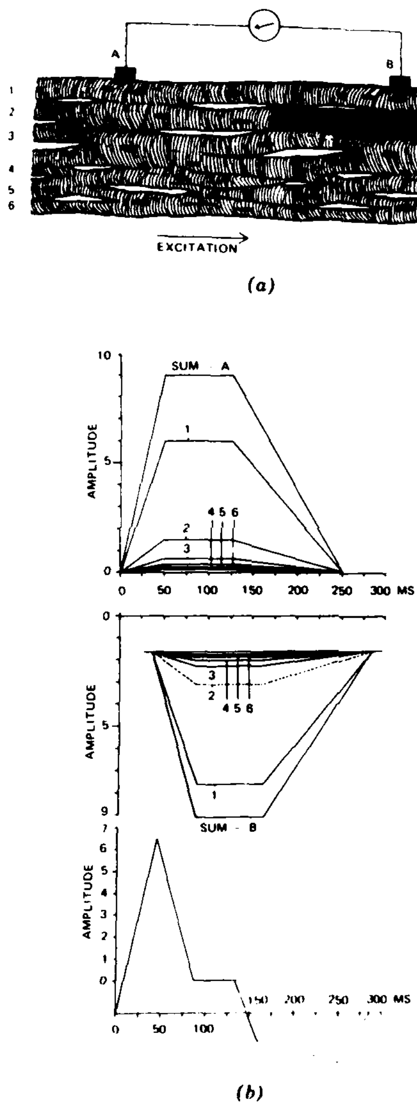


Figure 2. Schematic diagram of electromyographic recording. In part (a), two electrodes are shown positioned over six muscle fibers. In (b), the summed potential differences are shown for electrodes A and B, with the contributions from each fiber, and their difference. Reprinted from Geddes, 1972.

the wave of depolarization sweeps down the fibers and reaches the second electrode, it becomes negative. The event recorded also reflects the timing of the action potential passage at the two electrodes, and the size of the recording surface. In the example shown, there is a period when the fiber is depolarized under both electrodes; hence, the signal returns to zero before reversing its sign. Another factor determining the signal picked up by the electrodes is the intervening tissue. In general, the presence of tissue creates a low-pass filtering effect whose bandwidth decreases as distance increases (DeLuca, 1978).

While it is possible to record from a single muscle fiber (Ekstedt & Stålberg, 1973), the more usual recording represents events in a motor unit, or an aggregate of motor units. Under normal conditions, an action potential propagating down a motor nerve activates all the fibers of its motor unit. The fibers of a single motor unit are intermingled with each other in such a way that the territory of one unit is about 20 times the cross-sectional area of the fibers of the unit (Buchthal, Erminio, & Rosenfalk, 1959). Since a portion of a muscle might contain fibers belonging to any of fifty motor units, an electrode in the vicinity might detect activity in any or all of them. The signal reaching a pair of electrodes in active tissue is the weighted sum of the activity of each of the fibers of a motor unit, with the filtering properties of the tissue between the electrode and the active fiber taken into account. Since the orientation of the fibers of each motor unit with respect to a fixed recording site will be unique, the shape of the resulting recorded action potential will similarly be unique, and can be used to recognize the unit (LeFever, 1980).

When a muscle is activated, the electrical manifestation of a motor unit action potential is accompanied by a twitch of the activated fibers. In muscle contraction in physiological conditions, the motor units are repeatedly activated, whether the type of contraction is isometric (the muscle does not shorten, but develops tension) or anisometric (the muscle shortens).

THE ELECTRODE

In recordings from the laryngeal muscles, or any others, it is often possible to recognize individual motor units by visual inspection, especially when levels of contraction are low, so that only a few motor units are active. An example is shown in Figure 3, a recording from the cricothyroid muscle (Faaborg-Andersen, 1964). Alternatively, it is possible to record from such a large number of active fibers that individual components cannot be recognized, as in Figure 4. The signals shown here are a so-called "interference pattern." That is, the pattern represents the activity of a large number of fibers. The experimenter may wish to record single motor units or interference patterns, depending on the purpose of the experiment, and makes a choice of electrode accordingly.

Three general types of electrodes have been used in speech research; surface, needle, and hooked wire electrodes. Of these, hooked wire electrodes have been most useful for recording from the laryngeal muscles. The muscles of the larynx are aligned in a way that signals picked up by an electrode on the neck surface are ambiguous as to which muscle is the signal source. Thus,

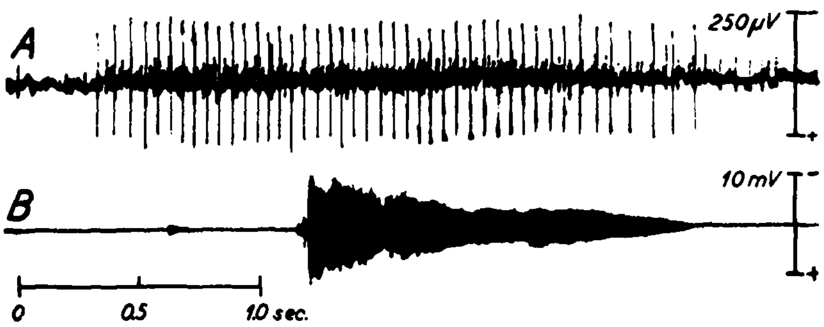


Figure 3. Action potentials of a single motor unit during phonation. A. Cricothyroid muscle. B. Microphone recording. Reprinted from D. Brewer, 1964.

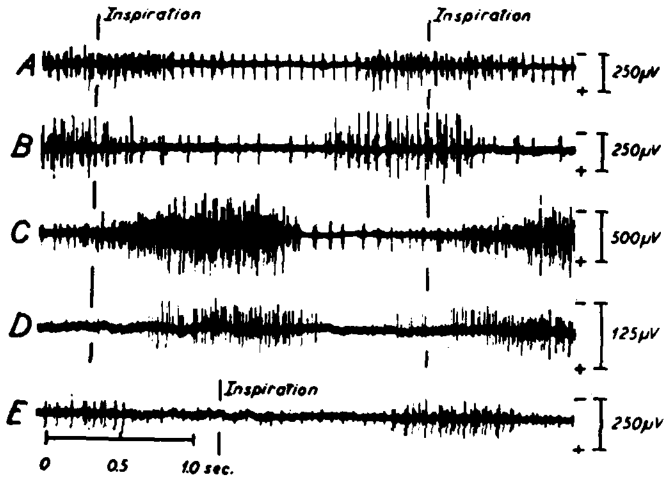


Figure 4. Quiet respiration. The onset of inspiration is indicated by the vertical stippled lines. A and B: Cricothyroid muscle. C and D: Vocalis muscle. E: Posterior cricoarytenoid muscle. Reprinted from D. Brewer, 1964.

although attempts have been made to use surface recordings from locations over the thyroid cartilage in a biofeedback application (Guitar, 1975), it seems unlikely that much further application will be made of such techniques. Needle electrode insertions into the laryngeal muscles are not generally feasible for posterior cricoarytenoid and interarytenoid muscles, although such insertions were used by Faaborg-Andersen in his classic study. The work of the past decade was done almost entirely with hooked wire electrodes, except for some clinical work to be described by Hirose.

Figure 5 shows the classic version of the hooked wire electrode (Basmajian & Stecko, 1962). Some technical details and possible variants of this type of electrode are discussed by Basmajian (1978). This type of electrode has been used in recording from the laryngeal muscles by a number of investigators besides ourselves (Hirano & Ohala, 1969; Shipp, Fishman, & Morrissey, 1970). Using them, we have been able to record from all of the intrinsic laryngeal muscles (and a wide variety of other speech muscles) using techniques developed collaboratively with Dr. Hajime Hirose and his colleagues at the Institute of Logopedics and Phoniatrics at the University of Tokyo (Hirose, Gay, & Strome, 1971).

If the investigator is interested in recording from a very small volume of tissue, the recording surfaces of the electrodes must be made as small as possible, while if the investigator is interested in a representation of the activity of the whole muscle, the recording surface must be as large as possible, while still remaining within the confines of the same muscle. Obviously, since the laryngeal muscles are small, some conventional configurations of electrode may record activity from more than one muscle (Dedo & Dunker, 1966). In the conventional hooked wire electrode, the hooks, which hold the wire in the muscle, also act as the recording points for the bipolar pickup, through their cut ends. However, the spacing between the two points is set arbitrarily by the way that the electrode happens to hook into the muscle, and, indeed, may change within the recording session (Jonsson & Komi, 1973). Since this type of electrode apparently records from a very small volume of tissue, the fact that the distance between the electrode tips is not fixed seems a design flaw. At Haskins, we have been exploring the various designs in which the functions of stabilization and recording are separated, and the field size is fixed by the separation between recording points.

PROPERTIES OF MOTOR UNITS

Exploring the relationship between ideal electrode and experiment requires a systematic discussion of the events within a muscle as we now know them, largely from studies of limb muscles. Most issues of muscle characteristics have only been explored with a limited number of muscles.

Let us begin with the single motor unit. In constant force contractions, it will fire with an overall mean interspike interval and standard deviation (DeLuca & Forrest, 1973; Figure 6), which can be used to characterize the unit, and, perhaps, the muscle itself. MacNeilage (1973) has shown that single motor units from CT and PCA fire at mean frequencies of about 15 impulses per second, during low frequency phonation. He suggested that these rates were intermediate between rates for limb and trunk and intraocular

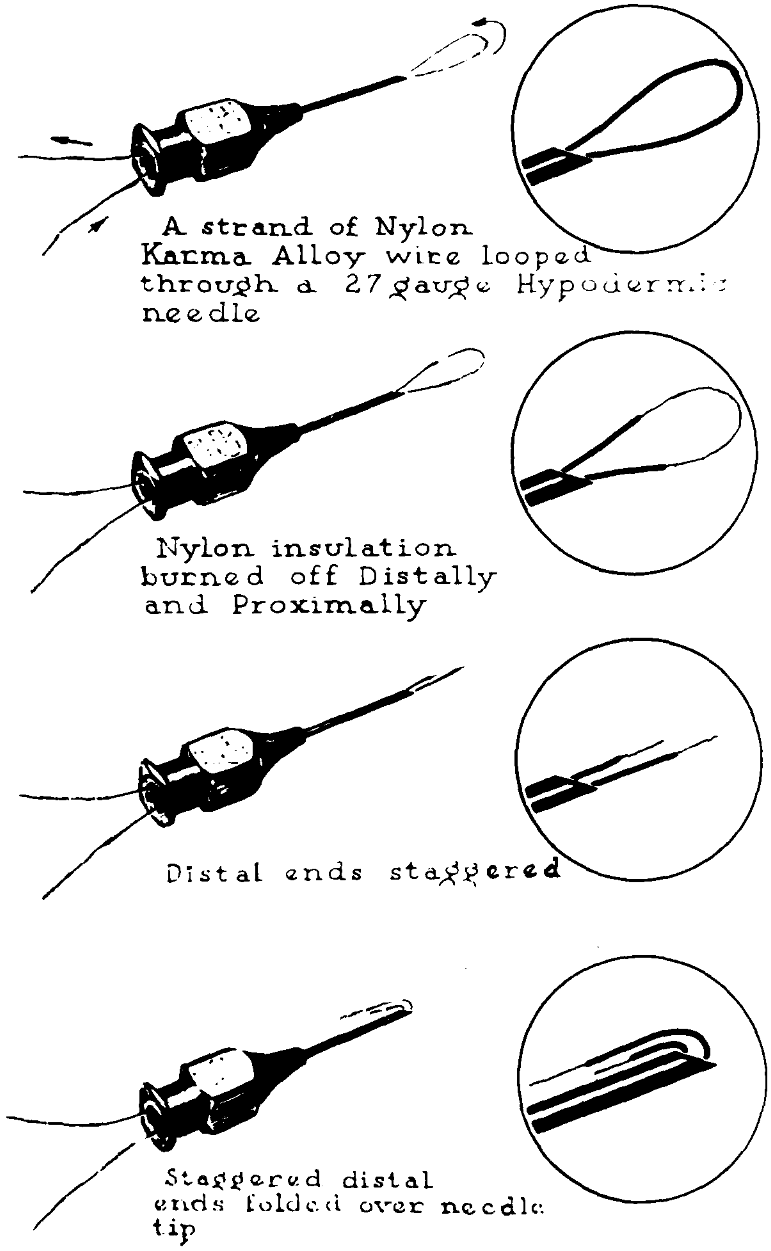


Figure 5. Steps in making a bipolar fine-wire electrode with the carrier needle used for insertion. Reprinted from Basmajian and Stecko, 1962.

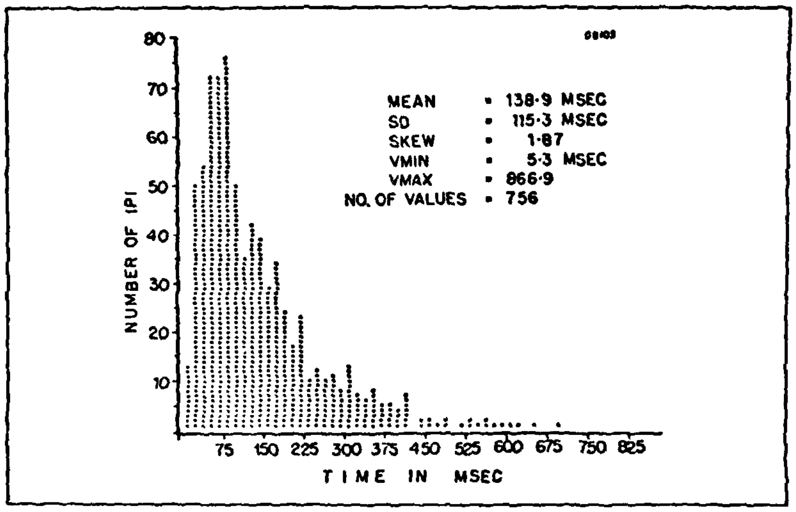


Figure 6. Distribution of interpulse intervals from a single motor unit. Reprinted from DeLuca and Forrest, 1973.

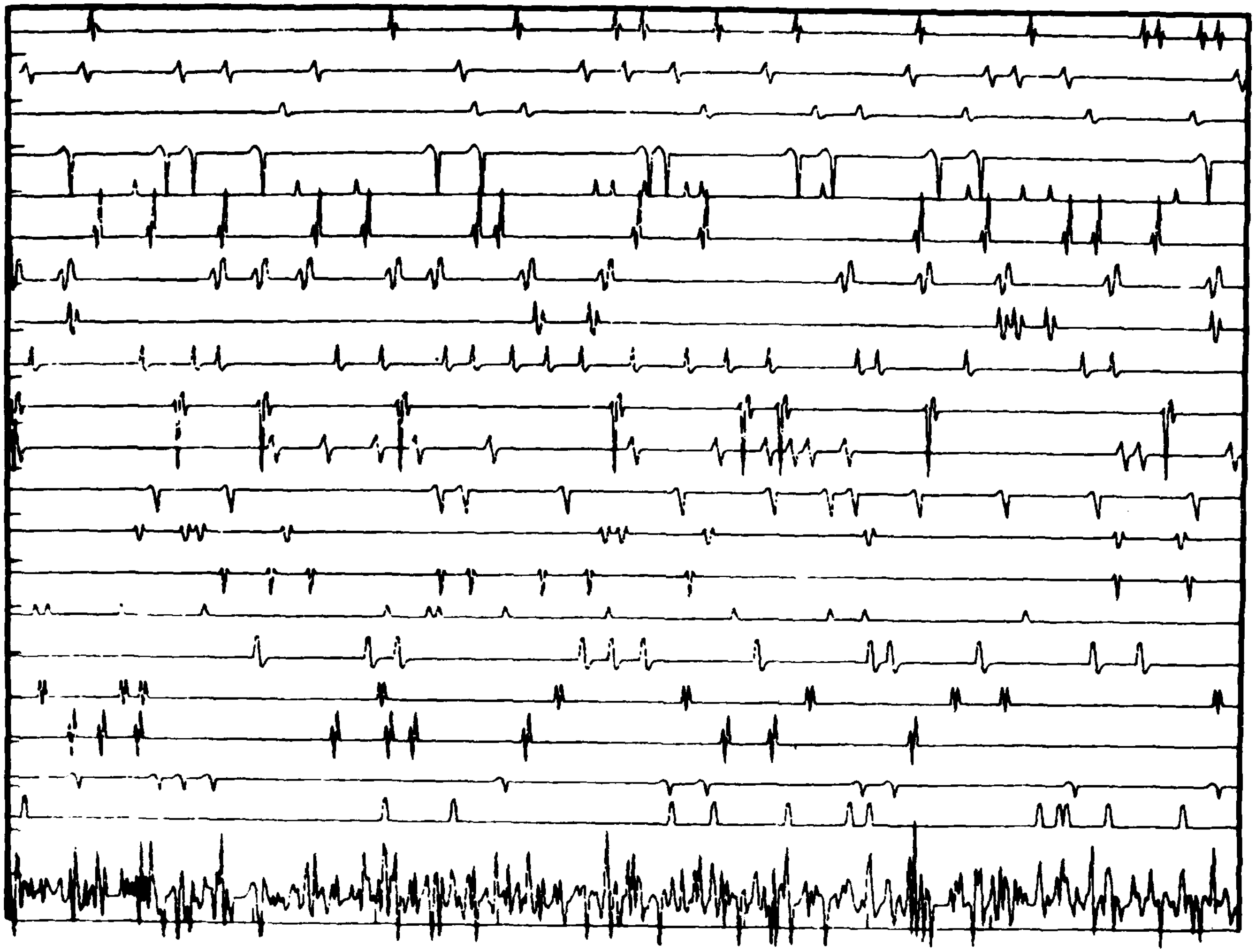


Figure 7. Synthetic interference pattern. The interference pattern at the bottom is the sum of the twenty "motor units" in the upper lines. C. DeLuca.

musculature, as we might expect from these other properties. However, he found no evidence for the different kinds of units, tonic and kinetic, postulated by Tokizane and Shimazu (1964), to be identifiable on the basis of the relationship between variability and firing rates (MacNeilage, Sussman, & Powers, 1977). Other authors (DeLuca & Forrest, 1973; Hannerz, 1974; Leifer, 1969) have found continuous distributions of single unit properties for various limb muscles.

During force-varying isometric contractions, there is a complex relationship between variation in firing rate and recruitment. At low forces, force tends to be increased by the recruitment of additional units, with successively recruited units having higher firing rates at recruitment. As force increases, individual units increase firing rates, and at the highest force levels, very little recruitment occurs. Synchronization of firing of units may occur as the muscle fatigues (DeLuca, 1978).

The most consistent observation of motor unit behavior is the relationship between the size of the unit, and force output and order of recruitment with increasing muscle force, the "size principle" (Henneman, 1975). While this relationship has not been observed for any of the laryngeal muscles, it has been demonstrated for the masseter in humans (Yemm, 1977) and for the anterior belly of the digastric by MacNeilage, Sussman, Westbury, and Powers (1979), and there is no reason to believe that the laryngeal muscles behave in a very unusual way in this respect. However, for all muscles, there is some question as to whether there are reversals of recruitment order for rapid, anisometric contractions.

Since the territories of motor units overlap with increasing forces of contraction, it is increasingly difficult to identify individual units. For studies of such questions, electrode size must be reduced, and sophisticated programs for the identification of motor units developed (LeFever, 1980).

THE INTERFERENCE PATTERN

Most electromyographic studies of the laryngeal muscles have been concerned, not with the properties of individual motor units, but with the functions of the muscles as a whole. Typically, the studies have related the characteristics of a given muscle activity to some sort of output, such as pitch. The electromyographic signal studied is usually an interference pattern, the signal from a large number of motor units. As an aid in visualization, it is interesting to look at a synthesized interference pattern, Figure 7 (LeFever & DeLuca, personal communication). The figure shows 20 motor units of shapes that would be characteristic of those found in an electrode field during a constant force, isometric contraction. Their sizes and the relative extent of positive and negative deviations from baseline vary with distance from and orientation to the electrode. The sum of positive and negative deviations is shown in the bottom line of the figure. Obviously, there is summing and cancellation of signals from individual units, depending on their phase relations. The resultant signal is noisy, and difficult to deal with quantitatively. If the electrode size is reduced, so that fewer units are represented in the signal, the interference pattern becomes more variable as a function of time (Figure 8A).

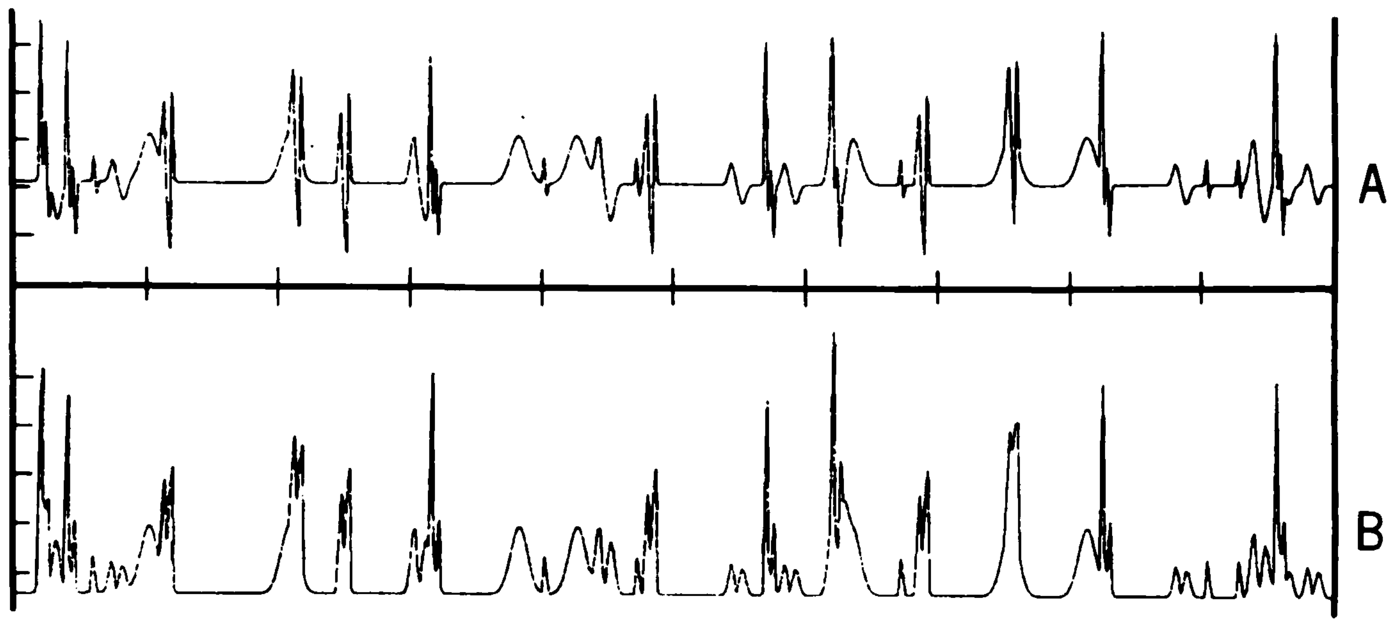


Figure 8. A. Synthetic interference pattern, sum of 5 motor units. B. The same interference pattern, after rectification. C. DeLuca.

A number of steps must be taken to deal with such signals. The usual approach has been to rectify and integrate. The effects of rectification are shown in Figure 8B. The traditional use of the rectified and integrated EMG signal is based on a large body of research investigating the relationship between the magnitude of the EMG signal so obtained and the force output of the muscle (Bigland & Lippold, 1954; Bouisset, 1973; Bouisset & Maton, 1973; Inman, Ralston, Saunders, Feinstein, & Wright, 1952; Lippold, 1952; Zuniga & Simons, 1969). This measure ("integrated EMG") varies roughly linearly with force for isometric contractions at moderate force levels, but at higher levels of force the relationsip becomes nonlinear. The situation becomes far more complex for anisometric contractions, in part because the mechanical efficiency of a muscle depends on its length as well as its velocity of shortening or lengthening. Since the events of interest in speech research are typically of this latter sort, we can expect the magnitude of the EMG signal to provide no more than an overall index of mechanical performance.

A possibility that we have explored informally at Haskins is calculating the variance of the interference pattern, which is equal to the sum of the variances of the motor unit action potential trains contributing, and hence, does not lead to the loss of contributions of motor units due to cancellation as does the more conventional measure.

We have said very little about the time constant to be used for integration. We use a 5 millisecond hardware integration window and smooth further algebraically, using software programs in which a time constant may be chosen. Individual tokens recorded with hooked-wire electrodes show sizable fluctuations that are not represented in the mechanical output of the muscle as a whole. For speech, time-smoothing is useful only to the point where it does not obscure the sequencing of underlying articulatory events. An alternative way of smoothing is ensemble averaging. The effects of time-smoothing and ensemble averaging are shown in Figure 9, which shows averaged and integrated signals from repeated utterances. The details of these analysis procedures are discussed at greater length in laboratory reports (Kewley-Port, 1973, 1974).

LARYNGEAL MUSCLE STUDIES

Having reviewed the general properties of muscles, and of the laryngeal muscles in particular, as well as some technical problems, we turn now to the results of electromyographic studies of the function of these muscles in speech. The most primitive question, is, perhaps, what muscles should be considered as laryngeal muscles? Traditionally, the muscles of the larynx have been divided into two groups, intrinsic and extrinsic. The identity of the intrinsic muscles is readily agreed upon; they are the cricothyroids (CT), the thyroarytenoids (TA), the interarytenoids (IA), the lateral cricoarytenoids (LCA), and the posterior cricoarytenoids (PCA). The identity of the extrinsic laryngeal muscles is more difficult to specify. If we take the empirical point of view that any muscle that affects the positions of thyroid, cricoid, and arytenoid cartilages relative to each other may be considered to be an extrinsic laryngeal muscle, then a wide variety of muscles, not normally considered in relation to the larynx, must be included. For example, Painter (1978) has produced some evidence that genioglossus activity may influence

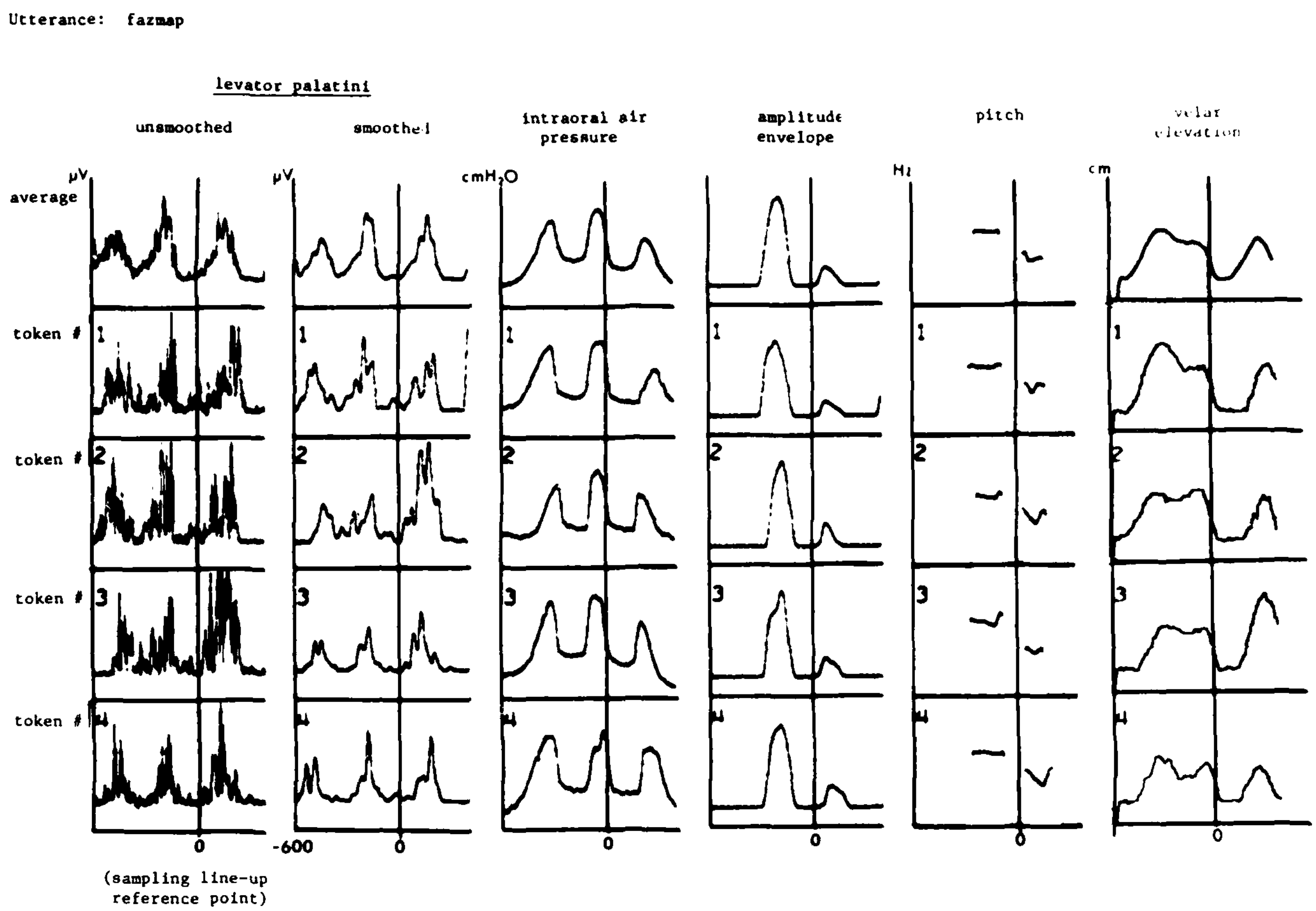


Figure 9. Individual and averaged tokens for the spoken utterance "faz map." The top row represents averages of 20 tokens. Four tokens are shown beneath the average. The first two columns show EMG output from the levator palatini, after sampling and rectification, before and after smoothing. The remaining columns show intraoral pressure, audio amplitude, fundamental frequency, and measured velar height. Haskins Laboratories.

pitch, and Erickson, Liberman, and Niimi (1977) have produced the same sort of evidence for geniohyoid. The implication is that a wide variety of muscles may affect pitch, as Sonninen suggested many years ago (1956). However, given the lack of detailed information about secondary effects on vocal fold adjustment, only the three strap muscles, the sternohyoid, the thyrohyoid, and the sternothyroid will be considered as extrinsics here.

Fundamental Frequency Control. Electromyographic studies on the regulation of pitch have been reported by many authors. More recent electromyographic studies have included those of Hirano, Vennard, and Ohala (1970), Shipp and McGlone (1971), Gay, Hirose, Strome, and Sawashima (1972), and Baer, Gay, and Niimi (1976).

These studies all conclude that cricothyroid activity increases as the pitch is raised, at least over most of the pitch range, as we might have expected from the mode of action of this muscle in producing torque around the cricothyroid joint. This action presumably underlies the observed lengthening of the folds with increasing f_o.

The activity of TA also increases as the pitch is raised over most of the pitch range, although it is more active in chest voice than in falsetto (Hirano, Ohala, & Vennard, 1969; Hirano et al., 1970; Baer et al., 1976), but the function of this activity is obscure. The thyroarytenoid could act, of course, to produce a shortening force in opposition to CT, although this cannot be its primary function, since its activity increases with pitch rise rather than pitch fall. One theory, by van den Berg (1960), as to its primary function suggests that it exerts "medial compression," limiting the horizontal extent of vocal fold vibration, permitting the more effective play of aerodynamic forces. An alternate possibility is that its tension is adjusted with compensating adjustments of CT, to tune the natural vibrating frequency of the muscle itself, considered as a tissue mass, since the muscle makes up the bulk of the folds and so determines, in large part, their vibratory characteristics. A secondary problem in the characterization of TA activity is that there is disagreement in the literature as to whether there are functional or anatomical differences between lateral and medial (vocalis) parts of TA, so that an adequate description of the function of one part may not suffice for the other (Sawashima, 1970).

Reports on the other laryngeal adductors, IA, LCA, and the more lateral parts of TA, tend to show increasing activity with increasing pitch. Van den Berg (1960) suggested, on the basis of cadaver experiments, that the IA might be active without the laterals at very low pitches, but this possibility has never been experimentally verified.

Some authors (e.g., Dedo, 1970; Gay et al., 1972; Baer et al., 1976) report increases of PCA activity at the highest f_o's when intensity is great, although there is not universal agreement on this point (Shipp & McGlone, 1971). Although this muscle is normally an abductor, its activity at high f_o is thought to brace the arytenoids against the anterior pull of the vocal folds. The observations of Gay et al. are summarized in Figure 10.

Control of f_o by the extrinsic muscles of the larynx is less well understood than control by the intrinsic muscles. The larynx, and f_o, move up

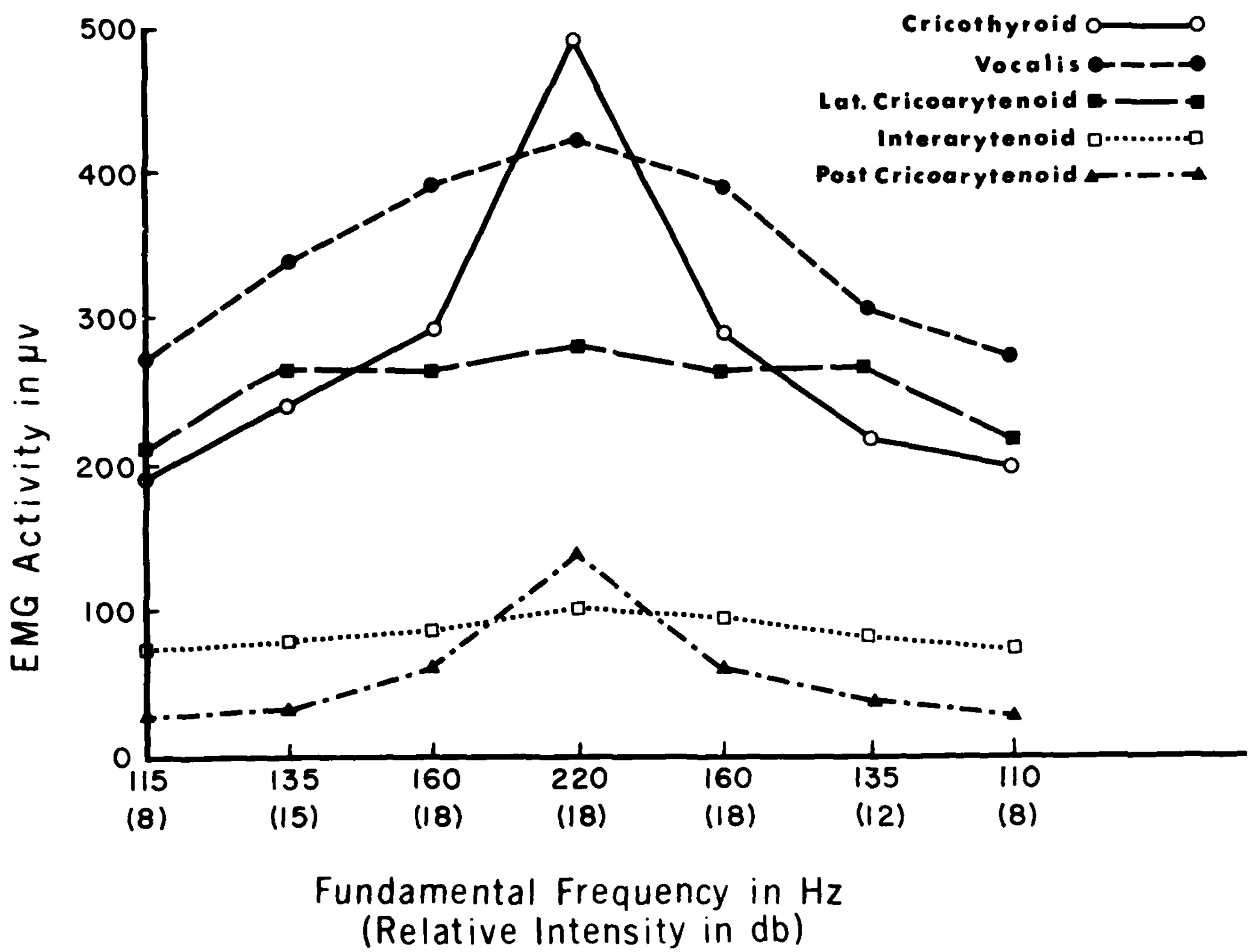


Figure 10. EMG activity for various laryngeal muscles as a function of frequency. From Gay, Hirose, Strome, and Sawashima, 1972.

and down during singing by untrained singers, or during speech, although trained singers learn to keep the larynx at an approximately constant low position (Sonninen, 1956; Shipp & Izdebski, 1975). These movements are produced largely by activity of the extrinsic attachments to the larynx, especially by the strap muscles.

Strap muscle activity (sternohyoid, sternothyroid) is correlated with f_o at both its highest and lowest levels. Although Kakita and Hiki (Note 1) have reported differentiation among these muscles, the weight of the evidence is that they act together in controlling pitch. This finding is supported both by electromyographic measurements (Faaborg-Andersen & Sonninen, 1960; Baer et al., 1976) and by clinical observation of patients who have had these muscles sectioned (Sonninen, 1956). Although, on anatomical grounds, it would seem that the sternothyroid muscle ought to increase f_o by tilting the thyroid cartilage down and forward, and that the thyrohyoid ought to decrease f_o by tilting the thyroid cartilage up and back, Sonninen showed that the situation is more complex. In experiments with cadavers and in stimulation experiments with patients undergoing thyroidectomy, he found that the effect on the larynx of activity of these muscles depended on posture and head position. The sternothyroid, in particular, can tilt the thyroid cartilage either way.

Sonninen developed an "external frame function" theory to account for f_o raising, based on his own results and those of other investigators. According to this theory, all the strap muscles work in conjunction with the anterior suprahyoid muscles. Although the strap muscles may or may not raise the larynx, their main function is to pull the thyroid cartilage forward. At the same time, activity of the cricopharyngeus and downward pull of the esophagus exert a downward and backward force on the posterior part of the cricoid cartilage.

Since the mechanism for application of the "external frame function" theory to f_o lowering has been elusive, alternative theories have been advanced. One of these is the passive theory, stating that f_o/larynx lowering is due to relaxation of the mechanisms for f_o/larynx raising. Although passive lowering can explain some of the observed relationships, two facts support the notion of at least an ancillary active mechanism. Electromyographic activity accompanies lowering as we noted above, and studies of vertical larynx position show that the position during low frequency phonation is lower than that in rest position (Shipp & Izdebski, 1975). A second theory, attributed to Ohala (1972), suggests that raising and lowering the larynx affects f_o directly through adjustment of the vertical tension of the vocal fold cover, which is continuous with the lining of the trachea. This theory cannot be adequately evaluated without improved understanding of the vibratory mechanism of the vocal folds and actual measurements of "vertical tension" in raised-larynx and lowered-larynx configurations. Finally, a theory accounting for f_o lowering by laryngealization has been proposed by Lindqvist (1969). This theory asserts that the vocal folds are shortened (and, incidentally, transglottal pressure is reduced) by activity of the muscle fibers of the aryepiglottic sphincter. This mechanism does not appear to require lowering of the larynx and hence does not explain the observed movements or associated EMG activity. It may operate jointly with or independently of other mechanisms.

Results of studies of strap muscle function in speech first suggested that although f_o falls were always accompanied by an increase in strap muscle activity, the activity did not always precede f_o falls, and showed substantial effects of segmental variables (Collier, 1975; Hirano et al., 1969). Later analysis, however, suggested that strap activity does precede pitch drops from a mid to low range (Atkinson & Erickson, 1977; Erickson et al., 1977).

A problem in studying pitch control in speech has been the difficulty of analyzing the relationships among f_o, subglottal pressure, and the antecedent activity of the large number of relevant muscles. One technique, which has been found useful, cross-correlates f_o and integrated EMG (Atkinson, 1978). The delay at which the correlation reaches a maximum can be used to estimate the response time of the muscle. The magnitude of the correlation at this delay can then be used in estimating the magnitude of that muscle's contribution to pitch control. The analysis can be further refined by dividing the fundamental frequency range into subranges. Atkinson's study shows the contribution of strap muscle activity to be greatest at low frequencies, while CT activity has its greatest effects at high frequencies. Although the data analyzed in the study were extremely limited, further exploitation of the technique seems warranted.

There is, nonetheless, a limit to the amount of reliance one can place on the results of gross correlation studies. An ingenious new technique for studying the relationship of f_o and the activity of the various laryngeal muscles has been suggested by Baer (1978). The technique was adapted from one originally designed for the study of skeletal muscles (Milner-Brown, Stein, & Yemm, 1973). Continuous records were made of electromyographic activity from laryngeal muscles and of voice fundamental frequency from a subject producing steady, sustained phonation at low f_o. The fundamental frequency record exhibits small perturbations around a nominally constant value. If we assume that these perturbations represent the response to the firing of single motor units in those muscles that control pitch, then an average-response computation of fundamental frequency triggered by single motor unit firing of any muscle should exhibit a systematic deviation in the interval immediately following the firings. Figure 11 shows the results of following this procedure for CT. Using this technique, muscles whose activity is grossly inter-correlated can be uncorrelated to examine their individual effects on some variable. We feel that this technique shows great promise in the application just suggested, and others.

Stricture Control and Voicing Features

A second dimension of laryngeal adjustment in speech is stricture control, the degree to which the laryngeal sphincter is closed by the approximation of the vocal folds. While these adjustments can be used to produce overall changes in voice quality, most speech studies of this dimension have been aimed at understanding the mechanism of consonant voicing.

Fiberoptic visualizations of the glottis (Sawashima, Abramson, Cooper, & Lisker, 1970; Kagaya, 1974) show that voiced and voiceless consonants are characterized by differences in glottal opening. It is the timing of the abduction and adduction of the folds, relative to the movement of the upper articulators, that distinguishes consonant classes within and across languages.

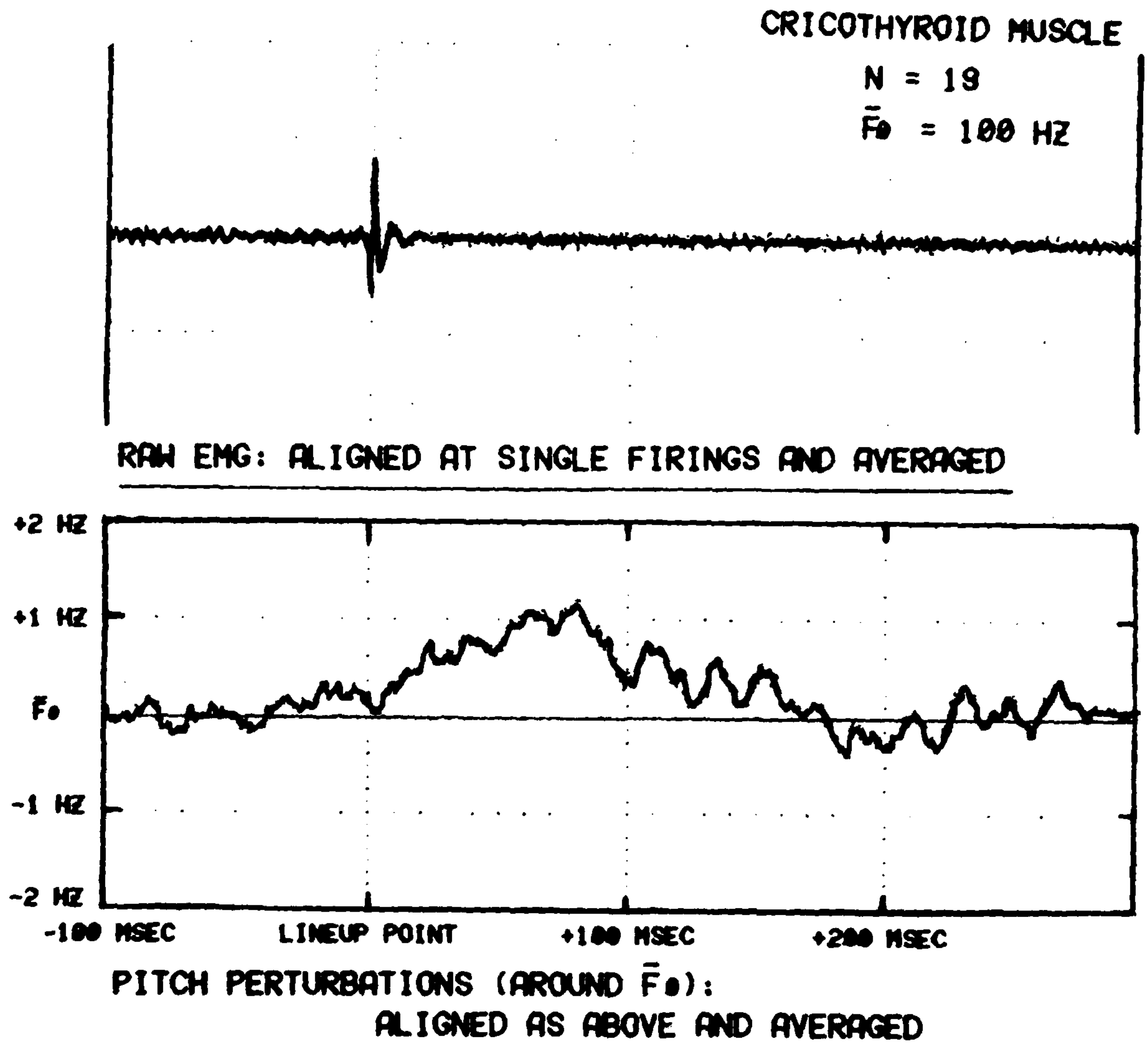


Figure 11. Single motor units of the cricothyroid, aligned and averaged, with parallel measure of pitch perturbation. See text for explanation. From Baer, 1981.

Anatomically, the five intrinsic laryngeal muscles can be divided into three functional groups with respect to stricture control: adductor (PCA), adductor (INT, TA, LAT), and tensor (CT). The question can then be asked whether the muscles function in speech in ways that the classification would suggest. Is there active abduction and adduction in voicing maneuvers? Do the adductors function together? Finally, is the activity of adduction and abduction accompanied by changes in tensing?

Abduction and adduction for voicing are clearly accomplished by the action of PCA and INT activity in a reciprocal way, as has been demonstrated in a number of studies (Hirose & Gay, 1973; Fischer-Jørgensen & Hirose, 1974; Hirose & Ushijima, 1976).

Figure 12 shows a fairly typical pattern obtained for this pair of muscles (Hirose, Lisker, & Abramson, 1972). The general conclusion is that the abductor (PCA) contracts, the adductor (INT) relaxes. The relationship has been quantified. Hirose (1977) showed that for a series of utterances containing voiced and voiceless stops, produced by a Japanese talker, the value of the correlation coefficient ranges between -.85 and -.65. The analysis does not make it clear what variables affect the value in a critical way.

The extent to which the activity of the adductor group is correlated in such maneuvers is still unclear. Some time ago, van den Berg and Tan (1959) showed, in cadaver experiments, that the different adductor muscles can be used to close the cartilagenous and membraneous parts of the larynx separately. Thus, we might expect some differences between the activity patterns of INT on the one hand, and LAT and TA on the other. Such differences have been seen in studies of Korean stops (Hirose, Lee, & Ushijima, 1974; Danish stød (Fischer-Jørgensen & Hirose, 1974) and glottal stops (Hirose & Gay, 1973). Apparently, the activity of LAT and TA is connected to the necessity for strong medial compression in these productions. However, the detail effects of differential contraction of these muscles on the shape of the glottis are not known. Figure 13 shows the contrast in activity of INT and VOC (TA) for the three types of voiceless stop found in Korean. The important point to note, apart from the obvious overall differences, is that there is a sharp peak in VOC activity for the glottalized Korean stop at consonant release, probably associated with increased tension of the folds.

A recent experiment by Yoshioka (1979) also suggests circumstances in which we perhaps will observe differentiation among laryngeal adductors in stricture control. He found that /h/ and /s/ may be produced with equal glottal widths, and equivalent patterns of reciprocal PCA and INT activity, but still differ in the presence of vibration at the edges of the memoranous portions of the folds in some examples of /h/. An obvious possibility is that other intrinsic laryngeal muscles show differences in activity for stricture control for the sounds.

A third question associated with the activity of the vocal folds in voicing control is whether activity of CT is associated with abduction or adduction. Stevens' model of glottal activity suggests that the tension of the vocal folds will affect the likelihood of vibration, for a given pressure drop across the glottis. It is therefore possible that some stops are

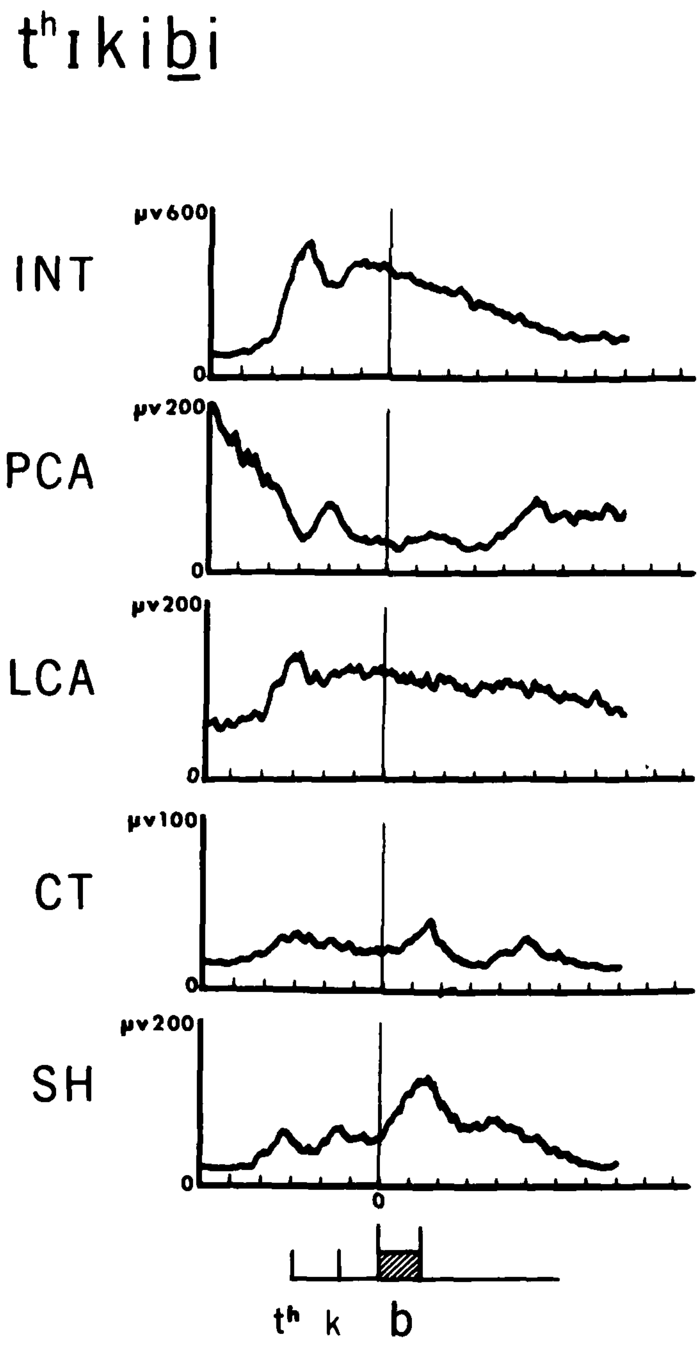


Figure 12. Intrinsic laryngeal muscle outputs for an utterance with a medial explosive voiced inaspirate [b]. The shaded interval at the bottom of the figure represents the duration of voicing during the [b] occlusion. From Hirose, Lisker, and Abramson, 1972.

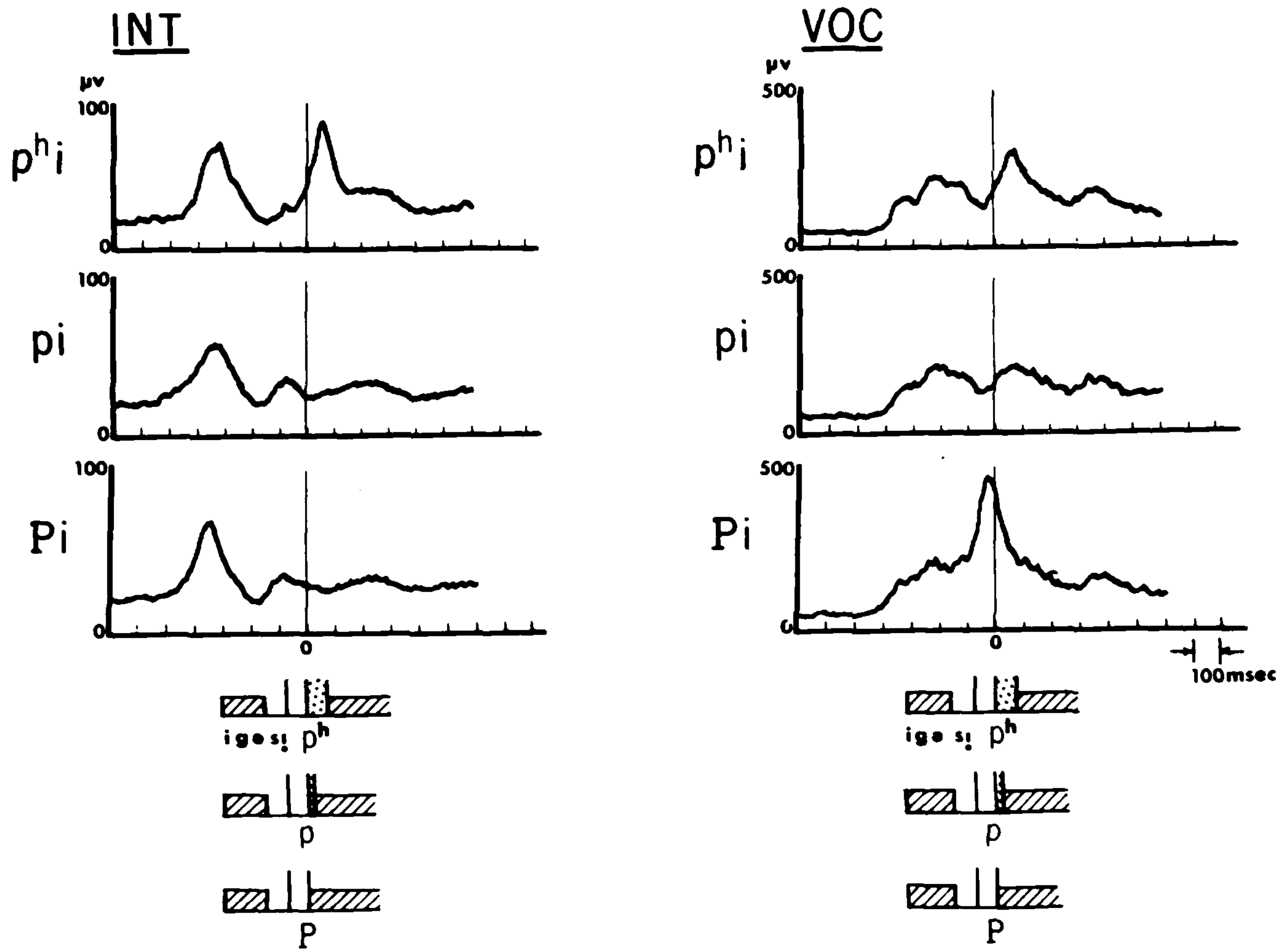


Figure 13. Averaged EMG curves for INT and VOC for the three bilabial stops of Korean. [$p^h i$] is voiceless and aspirate, [p] is voiceless and slightly aspirated, and [p] is voiceless and glottalized. From Hirose, Lee, and Ushijima, 1974.

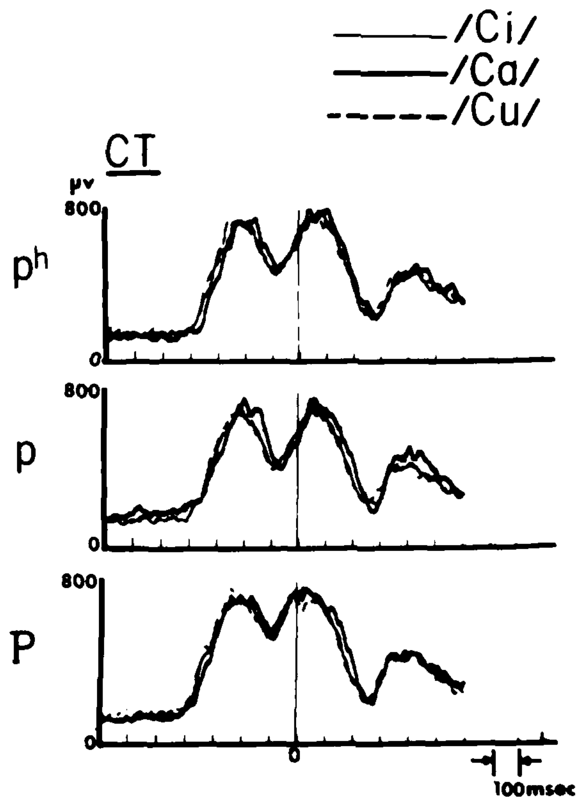


Figure 14. Cricothyroid activity for the three bilabial stops of Korean. The three curves in each box represent utterances containing the vowels /i/, /a/, and /u/. From Hirose, Lee, and Ushijima, 1974.

characterized by contrasts in CT activity, particularly those that contrast in degree of aspiration, like those of Korean (Hirose et al., 1974). A study of stop production in a single speaker (Figure 14) fails to support the hypotheses of CT differentiation, but small differences in CT activity accompanying voicing contrasts have been found from time to time.

The brief summary of laryngeal muscle function in this section and the preceding one reveal that we now have a gross qualitative sketch of the activity patterns, and the technical means at hand to elaborate this picture, to match models and observations of the larynx developed in other ways. However, we might now ask what clinical uses might be made of EMG using presently available techniques.

ELECTROMYOGRAPHY IN FUTURE DEVELOPMENTS

At present, EMG is widely used in diagnosis of neuromuscular disorders. It has not been used this way for the laryngeal muscles, although it perhaps could be. For example, it seems possible to detect abnormal single motor unit firing patterns in these muscles, abnormal synchronization of motor unit firings (Hirose, 1977), or, perhaps, to differentiate peripheral neurogenic and myogenic disorders.

Another use, from my point of view a very exciting one, is to use EMG as a technique for examining articulatory programming and its breakdown. The work described in this paper, and others, can be used to show a very tightly time-constrained coordination of laryngeal and supra-laryngeal events in running speech. Aspects of this coordination appear to break down in stuttering (Freeman & Ushijima, 1978), and in apraxia (Freeman, Sands, & Harris, 1978). While the broad perceptual consequences of breakdown in laryngeal coordination have often been described (e.g., Darley, Aronson, & Brown, 1975), it seems far more direct to look at the underlying failures of patterning. One of the most unfortunate consequences of the description of normal and abnormal speech in terms of transcriptional entitites has been to focus description of speech motor behavior on the attainment or failure of attainment of stationary acoustic or articulatory targets, rather than on the temporal prescription for coordinated activity. For normal speakers, we need to investigate what maintains these prescriptions, by systematically attempting to disrupt them. For abnormal speakers, we need, first, to describe the disrupted speech in terms of the constituent articulatory acts, and second, to investigate the relative roles of various factors, such as feedback, in maintenance of existing coordinations.

REFERENCE NOTE

1. Kakita, Y., & Hiki, S. A study of laryngeal control for voice pitch based on anatomical model. Paper presented at the Eight International Congress on Acoustics, London, July, 1974.

REFERENCES

Atkinson, J. E. Correlation analysis of the physiological factors controlling fundamental voice frequency. Journal of the Acoustical Society of America, 1978, 63, 211-222.

Atkinson, J. E., & Erickson, D. The function of strap muscles in speech: Pitch lowering or jaw opening? Haskins Laboratories Status Report on Speech Research, 1977, SR-49, 97-102.

Baer, T. Effect of single-motor-unit firings on fundamental frequency of phonation. Journal of the Acoustical Society of America, 1978, 64 (Suppl. 1), S90. (Abstract)

Baer, T. Investigation of the phonatory mechanism. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, this volume.

Baer, T., Gay, T., & Niimi, S. Control of fundamental frequency, intensity, and register of phonation. Haskins Laboratories Status Report on Speech Research, 1976, SR-45/46, 175-185.

Basmajian, J. Muscles alive (4th ed.). Baltimore: Williams and Wilkins, 1978.

Basmajian, J. V., & Stecko, G. A new bipolar indwelling electrode for electromyography. Journal of Applied Physiology, 1962, 17, 849.

Bigland, B., & Lippold, O. C. J. The relation between force, velocity and integrated electrical activity in human muscles. Journal of Physiology, 1954, 123, 214-224.

Bouisset, S. EMG and muscle force in normal motor activities. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 1). Basel: S. Karger, 1973, 547-583.

Bouisset, S., & Maton, B. Comparison between surface and intramuscular EMG during voluntary movement. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 1). Basel: S. Karger, 1973, 533-539.

Brandstater, M. E., & Lambert, E. H. Motor unit anatomy. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 1). Basel: S. Karger, 1973, 14-22.

Buchthal, F. Electromyography. In A. Remond (Ed.), Handbook of electromyography and clinical neurophysioloy (Vol. 16). Amsterdam: Elsevier, 1973.

Buchthal, F., Erminio, F., & Rosenfalk, P. Motor unit territory in different human muscles. Acta Physiologica Scandinavica, 1959, 45, 72-87.

Carlson, F. D., & Wilkie, D. R. Muscle physiology. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Collier, R. Physiological correlates of intonation patterns. Journal of the Acoustical Society of America, 1975, 58, 249-255.

Darley, F. L., Aronson, A. E., & Brown, J. R. Motor speech disorders. Philadelphia: W. B. Saunders Company, 1975.

Dedo, H. The paralyzed larynx. An electromyographic study of dogs and humans. Laryngoscope, 1970, 80, 1455-1517.

Dedo, H., & Dunker, E. The volume conduction of motor unit potentials. Electroencephalography and Clinical Neurophysiology, 1966, 20, 608-613.

DeLuca, C. Towards understanding the EMG signal. In J. Basmajian (Ed.), Muscles alive (4th ed.). Baltimore: Williams and Wilkins, 1978, 53-78.

DeLuca, C. J., & Forrest, W. J. Properties of motor unit action potential trains. Kybernetics, 1973, 12, 160-168.

Edström, L., & Kugelberg, E. Histochemical composition, distribution of fibers and fatiguability of single motor units in the anterior tibial muscle of the rat. Journal of Neurology, Neurosurgery and Psychiatry, 1968, 31, 424-433.

Edström, L., Lindquist, C., & Mårtensson, A. Correlation between funcational and histochemical composition of the laryngeal muscles in the cat. In B. Wyke (Ed.), Ventilatory and phonatory control systems. London: Oxford University Press, 1974, 392-403.

Ekstedt, J., & Stålberg, E. Single fibre electromyography for the study of the microphysiology of the human muscle. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 1). Basel: S. Karger, 1973, 89-112.

English, D. T., & Blevins, C. E. Motor units of laryngeal muscles. Archives of Otolaryngology, 1969, 89, 782-785.

Erickson, D., Liberman, M., & Niimi, S. The geniohyoid and the role of the strap muscles. Haskins Laboratories Status Report on Speech Research, 1977, SR-49, 103-110.

Faaborg-Andersen, K. Electromyographic investigation of intrinsic laryngeal muscles in humans. Acta Physiologica Scandinavica, 1957, 41, Suppl. 140, 1-149.

Faaborg-Andersen, K. L. Electromyography of the laryngeal muscles in man. In D. W. Brewer (Ed.), Research potentials in voice physiology. New York: State University of New York, 1964, 105-129.

Faaborg-Anderson, K., & Sonninen, A. The function of the extrinsic laryngeal muscles at different pitch. Acta Oto-laryngologica, 1960, 51, 89-93.

Fischer-Jørgensen, E., & Hirose, H. A note on laryngeal activity in the Danish "stød." Haskins Laboratories Status Report on Speech Research, 1974, SR-39/40, 255-259.

Freeman, F. J., Sands, E. S., & Harris, K. S. Temporal coordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. Brain and Language, 1978, 6, 106-111.

Freeman, F. J., & Ushijima, T. Laryngeal muscle activity during stuttering. Journal of Speech and Hearing Research, 1978, 21, 538-562.

Gay, T., Hirose, H., Strome, M., & Sawashima, M. Electromyography of the intrinsic laryngeal muscles during phonation. Annals of Otology, Rhinology and Laryngology, 1972, 81, 401-409.

Geddes, L. A. Electrodes and the measurement of bioelectric events. New York: Wiley, 1972.

Guitar, B. Reduction of stuttering frequency using analog electromyographic feedback. Journal of Speech and Hearing Research, 1975, 18, 672-685.

Hannerz, J. Discharge properties of motor units in relation to recruitment order in voluntary contraction. Acta Physiologica Scandinavica, 1974, 91, 374-384.

Hast, M. H. Studies of the extrinsic laryngeal muscles. Archives of Otolaryngology, 1968, 88, 273-278.

Hast, M. H. The primate larynx. A comparative physiological study of intrinsic muscles. Acta Otolaryngologica, 1969, 67, 84-92.

Henneman, E. Principles governing distribution of sensory input to motor neurons. In E. Evarts (Ed.), Central processing of sensory input leading to motor output. Cambridge, Mass.: MIT Press, 1975, 281-293.

Hirano, M., & Ohala, J. Use of hooked-wire electrodes for electromyogaphy of the intrinsic laryngeal muscles. Journal of Speech and Hearing Research, 1969, 12, 362-373.

Hirano, M., Ohala, J., & Vennard, W. The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. Journal of Speech and Hearing Research, 1969, 12, 616-628.
Hirano, M., Vennard, W., & Ohala, J. Regulation of register, pitch, and intensity of voice. Folia Phoniatrica, 1970, 22, 1-20.
Hirose, H. Electromyography of the larynx and other speech organs. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977.
Hirose, H., & Gay, T. Laryngeal control in vocal attack. An electromyographic study. Folia Phoniatrica, 1973, 25, 203-213.
Hirose, H., Gay, T., & Strome, M. Electrode insertion techniques for laryngeal electromyography. Journal of the Acoustical Society of America, 1971, 50, 1449-1450.
Hirose, H., Lee, C. Y., & Ushijima, T. Laryngeal control in Korean stop production. Journal of Phonetics, 1974, 2, 145-152.
Hirose, H., Lisker, L., & Abramson, A. S. Physiological aspects of certain laryngeal features in stop production. Haskins Laboratories Status Report on Speech Research, 1972, SR-31/32, 183-191.
Hirose, H., & Ushijima, T. More on laryngeal control for voicing distinction in Japanese consonant production. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1976, 10, 101-112.
Hirose, H., Ushijima, T., Kobayashi, T., & Sawashima, M. An experimental study of the contraction properties of the laryngeal muscles in the cat. Annals of Otolaryngology, 1969, 78, 297-307.
Inman, V. T., Ralston, H. J., Saunder, J. B., Feinstein, B., & Wright, E. W. Relation of human electromyogram to muscular tension. Electroencephalography and Clinical Neurophysiology, 1952, 4, 187-194.
Jonsson, B., & Komi, P. V. Reproducibility problems when using wire electrodes in electromyographic kinesiology. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 1). Basel: S. Karger, 1973, 540-546.
Kagaya, R. A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics, 1974, 2, 161-180.
Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research, 1973, SR-33, 173-184.
Kewley-Port, D. An experimental evaluation of the EMG data processing system: Time constant choice for digital integration. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 65-72.
Kugelberg, E. Properties of the rat hind-limb motor units. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 1). Basel: S. Karger, 1973, 2-13.
LeFever, R. Statistical analysis of concurrently active human motor units. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1980.
Leifer, L. J. Characterization of single muscle fiber discharge during voluntary isometric contraction of biceps brachii muscle in man. Unpublished doctoral dissertation, Stanford University, 1969.
Liddell, E. G. T., & Sherrington, C. S. Further observations on myotatic reflexes. Proceedings of the Royal Society B, 1925, 97, 267-283.
Lindqvist, J. Laryngeal mechanisms in speech. Quarterly Progress and Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), 1969, STL-QPSR 2-3, 26-32.

Lippold, O. C. J. The relation between integrated action potentials in a human muscle and its isometric tensions. Journal of Physiology, 1952, 117, 492-499.
MacNeilage, P. F. Preliminaries to a study of single motor unit activity in speech musculature. Journal of Phonetics, 1973, 1, 55-71.
MacNeilage, P. F., Sussman, H. M., & Powers, R. K. Activation of motor units in speech musculature. Journal of Phonetics, 1977, 5, 135-148.
MacNeilage, P. F., Sussman, H. M., Westbury, J. R., & Powers, R. K. Mechanical properties of single motor units in speech musculature. Journal of the Acoustical Society of America, 1979, 65, 1047-1052.
Mårtensson, A., & Skoglund, C. R. Contraction properties of the intrinsic laryngeal muscles. Acta Physiologica Scandinavica, 1964, 60, 318-336.
Milner-Brown, H. S., Stein, R. B., & Yemm, R. The contractile properties of human motor units during voluntary isometric contractions. Journal of Physiology, 1973, 228, 285-306.
Ohala, J. How is pitch lowered? Journal of the Acoustical Society of America, 1972, 52, 124. (Abstract)
Painter, C. Implosives, inherent pitch, tonogenesis and laryngeal mechanisms. Journal of Phonetics, 1978, 6, 249-274.
Person, R. S., & Kudina, L. P. Discharge frequency and discharge pattern of human motor units during voluntary contraction of muscles. Electroencephalography and Clinical Neurophysiology, 1972, 32, 471-483.
Sahgal, V., & Hast, M. H. Histochemistry of primate laryngeal muscles. Acta Otolaryngologica, 1974, 78, 277-281.
Sawashima, M. Laryngeal research in experimental phonetics. Haskins Laboratories Status Report on Speech Research, 1970, SR-23, 69-115. [Also in T. Sebeok (Ed.), Current trends in linguistics (Vol. 12). The Hague: Mouton, 1976].
Sawashima, M., Abramson, A. S., Cooper, F. S., & Lisker, L. Observing laryngeal adjustments during running speech. Phonetica, 1970, 22, 193-201.
Shipp, T., Fishman, B. V., & Morrissey, P. Method and control of laryngeal EMG electrode placement in man. Journal of the Acoustical Society of America, 1970, 48, 429-430.
Shipp, T., & Izdebski, K. Vocal frequency and vertical larynx positioning by singers and nonsingers. Journal of the Acoustical Society of America, 1975, 58, 1104-1106.
Shipp, T., & McGlone, R. E. Laryngeal dynamics associated with voice frequency change. Journal of Speech and Hearing Research, 1971, 14, 761-768.
Sonninen, A. The role of the external laryngeal muscles in length adjustment of the vocal cords in singing. Acta Oto-laryngologica, 1956, Suppl. 130.
Tokizane, T., & Shimazu, H. Functional differentiation of human skeletal muscle. Springfield, Ill.: Thomas, 1964.
van den Berg, J. Myoelastic-aerodynamic theory of voice production. Journal of Speech and Hearing Research, 1958, 1, 227-244.
van den Berg, J. Vocal ligaments versus registers. In Current problems in phoniatrics and logopedics. Basel: S. Karger, 1960, 1, 19-34.
van den Berg, J., & Tan, T. S. Results of experiments with human larynxes. Practica Oto-Rhino-Laryngologica, 1959, 21, 425-450.
Yemm, R. The orderly recruitment of motor units of the masseter and temporalis muscles during voluntary isometric contraction in man. Journal of Physiology (London), 1977, 265, 163-174.

Yoshioka, H. Laryngeal adjustments during Japanese fricative and devoiced vowel production. Haskins Laboratories Status Report on Speech Research, 1979, SR-58, 147-160.

Zuniga, E. N., & Simons, D. G. Nonlinear relationship between averaged electromyogram potential and muscle tension in normal subjects. Archives of Physical Medicine and Rehabilitation, 1969, 50, 613-620.

INVESTIGATION OF THE PHONATORY MECHANISM*

Thomas Baer

Abstract. A rational approach toward the development of improved techniques for the prevention, detection, diagnosis, and correction of vocal pathologies rests on an improved understanding of voice mechanisms. To achieve these goals, we need to better understand the dimensions of phonatory performance and their dependence both on the state of laryngeal structures and on patterns of control. Because of the inaccessible location of the larynx, few direct measurements of this performance are possible. Quantitative mathematical modeling is a useful vehicle for studying laryngeal vocal function. Continuation and extension of excised-larynx and animal studies can provide detailed data in support of the development and testing of these models. Human experiments, *in vivo*, aimed at factoring out the phonatory consequences of variations in individual laryngeal control parameters are suggested as a means of further extending such studies.

INTRODUCTION

A rational approach toward the development of improved techniques for the prevention, detection, diagnosis, and correction of vocal pathologies rests on an improved understanding of voice mechanisms. For prevention, we hope to understand the pattern of control, and its correlates in vibratory performance, whose breakdown leads to physiological failures in the laryngeal structures. Our research in detection and diagnosis is directed toward isolating non-invasive multidimensional measures capable of differentiating performance of larynges with different pathologies from the performance of normal larynges and from each other. In the area of correction, we hope to improve the conceptual framework for voice training and therapy, and improve the ability of surgeons to predict the phonatory consequences of alternative procedures. To achieve these goals, we need to better understand the dimensions of phonatory performance and their dependence both on the state of laryngeal structures and on patterns of control.

The process of phonation can be separated into three components: a phonatory system, its inputs, and its outputs. The system consists of two subsystems: one aerodynamic (the glottis), and the other mechanical (the

---

*A version of this paper was presented at the Conference on Assessment of Vocal Pathology, Bethesda, Md., April 1979. (Proceedings to be published in *ASHA Reports*.)
Acknowledgement. This work was supported by NINCDS Grant NS13870 and BRS Grant RR05596.

vocal folds). Inputs to this system are muscular adjustments, transglottal pressure, and some other less significant variables. Ouputs may be considered to be the pattern of mechanical vibrations in the vocal folds, or, more significantly for voice production, the pattern of airflow into the vocal tract. This latter output then serves as input to another system--the vocal tract--whose output is the radiated voice signal.

The myoelastic-aerodynamic theory of phonation (van den Berg, 1958) accounts grossly for the nature of phonation in terms of a passive interaction between the two phonatory subsystems when an appropriate combination of inputs is applied. The acoustic theory of speech (Fant, 1960) accounts for the effects of the vocal tract in transforming the glottal source signal to a radiated acoustic output signal. Although both of these theories have been well known for two decades or more, there are significant details that remain poorly understood. Thus, we have only limited ability to estimate the glottal volume velocity waveform by canceling the effects of the vocal tract from the speech output signal, and we have only limited ability to separate the influences of inputs to the phonatory system from the influences of the system itself on details of its output. Because of the inaccessible location of the larynx, few direct measurements of this output are possible.

Investigations into the mechanisms of phonation and its control have relied heavily on research with models. Much basic knowledge can be derived from experiments with excised larynges (e.g., van den Berg & Tan, 1959) and with live animal preparations, which serve as simplified models of their intact counterparts but which can be more carefully observed and more systematically controlled. Fabricated mechanical models have also been used to test hypotheses about the mechanism. For example, Smith (1962) experimented with a "membrane-cushion" model, which seems to incorporate some elements of the more recent "cover-body" theory of Hirano (1974, 1975, 1977). Mostly, however, mathematical descriptions and computer simulations have been used to formalize and refine knowledge about the mechanisms. Thus, the development of these models is both a goal and a tool of phonatory research.

The history of these modeling efforts parallels the improvement of our understanding of the system. As our understanding has become more complete, the models have become more complex. Building on the aerodynamic studies of van den Berg, Zantema, and Doornenbal (1957), Flanagan and Landgraf (1968) modeled the vocal folds as a simple mass-spring system performing horizontal movements with one degree of freedom. It soon became apparent that an additional degree of freedom was required to account for vertical phase differences. Ishizaka and Matsudaira (1972) corrected some errors in van den Berg's aerodynamic analysis, and showed that a two-mass model of the vocal folds could more realistically account for the conditions under which phonation could be initiated. Ishizaka and Flanagan (1972) simulated the two-mass model, extending the results of Ishizaka and Matsudaira, but were limited by this model's inability to account realistically for the closed period of the glottal cycle. Titze (1973, 1974) increased the number of masses to 16, in order to allow a distribution of vibrations along the anterior-posterior direction. This model also allowed for some vertical movements. Finally, Titze and Talkin (1979) have been investigating more sophisticated models that explicitly model the layered structure of the vocal folds (Hirano, 1974) and their behavior as a vibrator, and that incorporate tissue viscosity and bulk incompressibility.

Though it is understood that models must be complex to account realistically for the phonatory mechanism, there is also a danger inherent in the growth of complexity. As the number of degrees of freedom and the number of independent parameters multiply, the possibilities for accurately modeling the detailed mechanism improve, but so do the possibilities for producing apparently realistic behavior due to mechanisms that may not represent those of the real larynx. For our purposes, models must be mechanistically correct as well as descriptive of the output. It is therefore essential to determine as many of their parameters as possible and the constraints among them by direct measurement, and to evaluate the performance of these models in the greatest possible detail. Furthermore, we ought to be able to make directly testable predictions on the basis of our modeling efforts.

Further progress in understanding the detailed mechanism of phonation and in developing an accurate model of it thus depends on detailing the mechanical characterisitics of vocal folds and determining their variation as functions of laryngeal control. It also depends on improved methods for measuring more detailed performance characteristics of real larynges, for comparing model performance to the performance of real larynges, and for generating testable predictions from modeling studies. Hirano has discussed, both at the *Conference on Assessment of Vocal Pathology* and in other publications (Hirano, 1975, 1977), measurements of mechanical properties of the vocal folds and some patterns of their variation with the contractions of individual muscles. Other papers at the conference will discuss techniques for obtaining detailed measurements, and Titze's paper will discuss methods for comparing the performance of models with these measurements on *in vivo* larynges. In the remainder of this paper, the continuation and extension of excised larynx and animal studies is urged because of their ability to produce detailed data for the direct testing of models. Then, some experiments *in vivo*, aimed at factoring out the phonatory consequences of variations in individual control parameters, are suggested as a means of further extending these studies.

I. EXPERIMENTS WITH EXCISED LARYNGES AND ANIMALS

It is well known that excised larynges, both canine and human, can simulate many of the vibratory characteristics of normal human larynges when they are attached to a pseudosubglottal system that supplies suitably conditioned airflow and when the positions of the laryngeal cartilages are suitably controlled, using strings to simulate the functions of muscles. As a simplified model of their intact counterparts, excised larynges offer several advantages. Because they are more accessible, they can supply observations and measurements that cannot be made *in vivo*. For example, both Matsushita (1969) and Baer (1975) have developed techniques for observing vibration patterns both from the normal supraglottal aspect and from the subglottal aspect. Baer also developed a technique for marking the vocal folds with small particles and tracking their frontal-plane movement trajectories throughout a glottal cycle using a microscope and stroboscopic illumination. Measurements could be made from both the supraglottal and subglottal aspects, and with the aid of qualitative observations, vocal fold shapes in the frontal plane throughout a cycle could be reconstructed from the measurements. With excised larynges, measurements of subglottal pressure and glottal airflow can be simplified. Furthermore, almost any technique for measuring characteris-

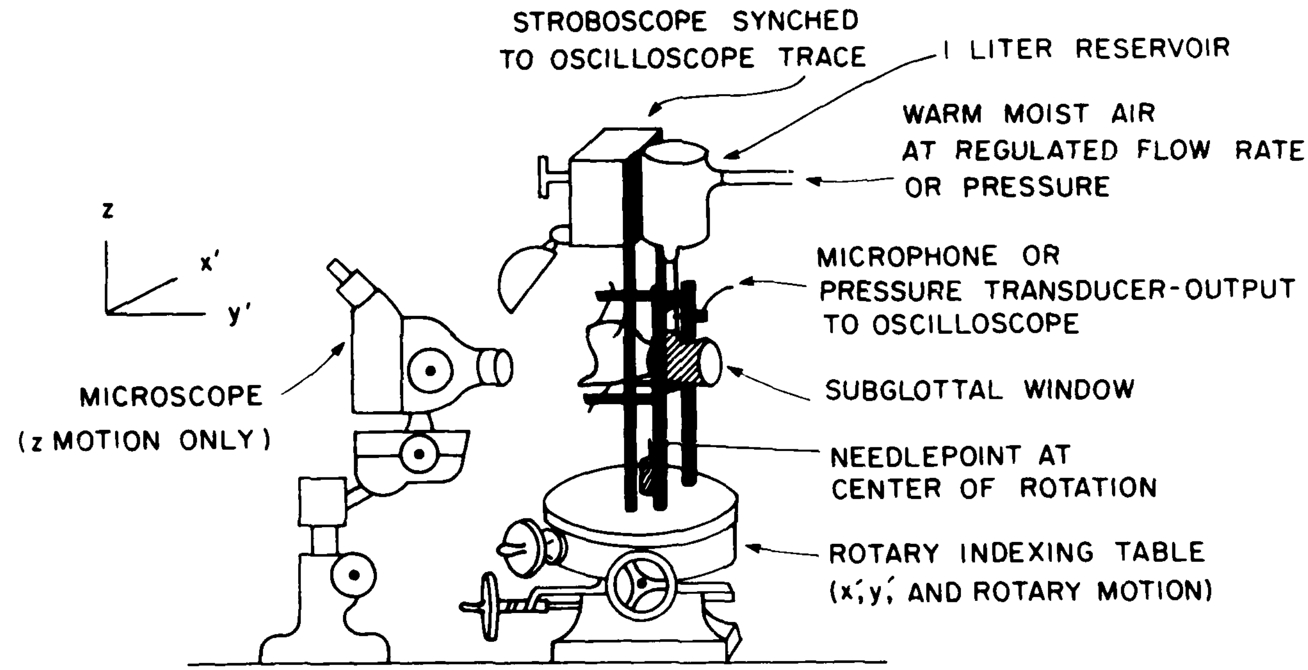


Figure 1. Schematic diagram of apparatus for measuring vibration patterns of excised larynges.

tics of phonatory vibrations can be used more effectively on an isolated larynx. Additional advantages are that the configuration of an excised larynx can be held constant or systematically varied, that its structures can be experimentally modified to determine the effects on vibration, and that they are accessible for measurement of mechanical properties in their configuration for voice production. The major limitations of the excised preparation--namely, that its death changes some of its mechanical properties, including its ability to tense the vocalis muscle--can be overcome by using live animal preparations and stimulating the muscles electrically. However, these advantages have not been fully exploited.

Baer's work with excised larynges was directed toward elucidating the phonatory mechanism in excised canine larynges. Although there is not space here to describe these experiments in detail, some of the most significant results are summarized below.

The experimental apparatus is shown schematically in Figure 1. A larynx was mounted on a pseudo-trachea, which made a right-angle turn just below the larynx, allowing a window to obtain a subglottal view. A stroboscope synchronized to subglottal pressure variations was mounted in front of the preparation. The phase at which the stroboscope was triggered could be adjusted to any point within the glottal cycle. Airflow was delivered at regulated flow rate or pressure, and both average pressure and average flow rate were measured. The subglottal system was intended to simulate the acoustic properties of the real subglottal tract. The apparatus was mounted on the top of a rotary indexing table, whose tabletop could be rotated, so that observations could be made through the microscope at any angle. The tabletop could also be translated along its two horizontal axes. A measurement system was devised by which the locations of any points observed through the microscope could be determined in three dimensions.

With respect to gross aspects of the performance of excised larynges, observations already made by others were replicated. In addition, it was observed that, for a given laryngeal configuration, phonation could be maintained at values of subglottal pressure below those required for initiating phonation. As the tissues desiccated, the separation between conditions for onset and conditions for maintenance increased. Thus, mobility of the surface tissues appeared to be important for initiating phonatory vibration. Perhaps this observation has some implications for the assessment of pathologies.

Figure 2 shows data from a run in which the frontal-plane trajectories of three particles were measured at eighth-cycle increments while the larynx sustained steady-state vibration. One particle was on the lateral superior surface of the vocal folds, a second was near the medial superior surface of the folds, and a third was on the lower (subglottal) surface. These trajectories are typical. They were roughly elliptical, in the clockwise direction (for the coordinate system shown). The minor axis of the ellipses decreased as average distance from the midline increased. Subglottal particles moved primarily in a horizontal direction, while supraglottal particles well off the midline moved primarily in a vertical direction. Trajectories of particles near the midline often exhibited complex perturbations near the superior-medial parts of their trajectories. Trajectories of the two upper

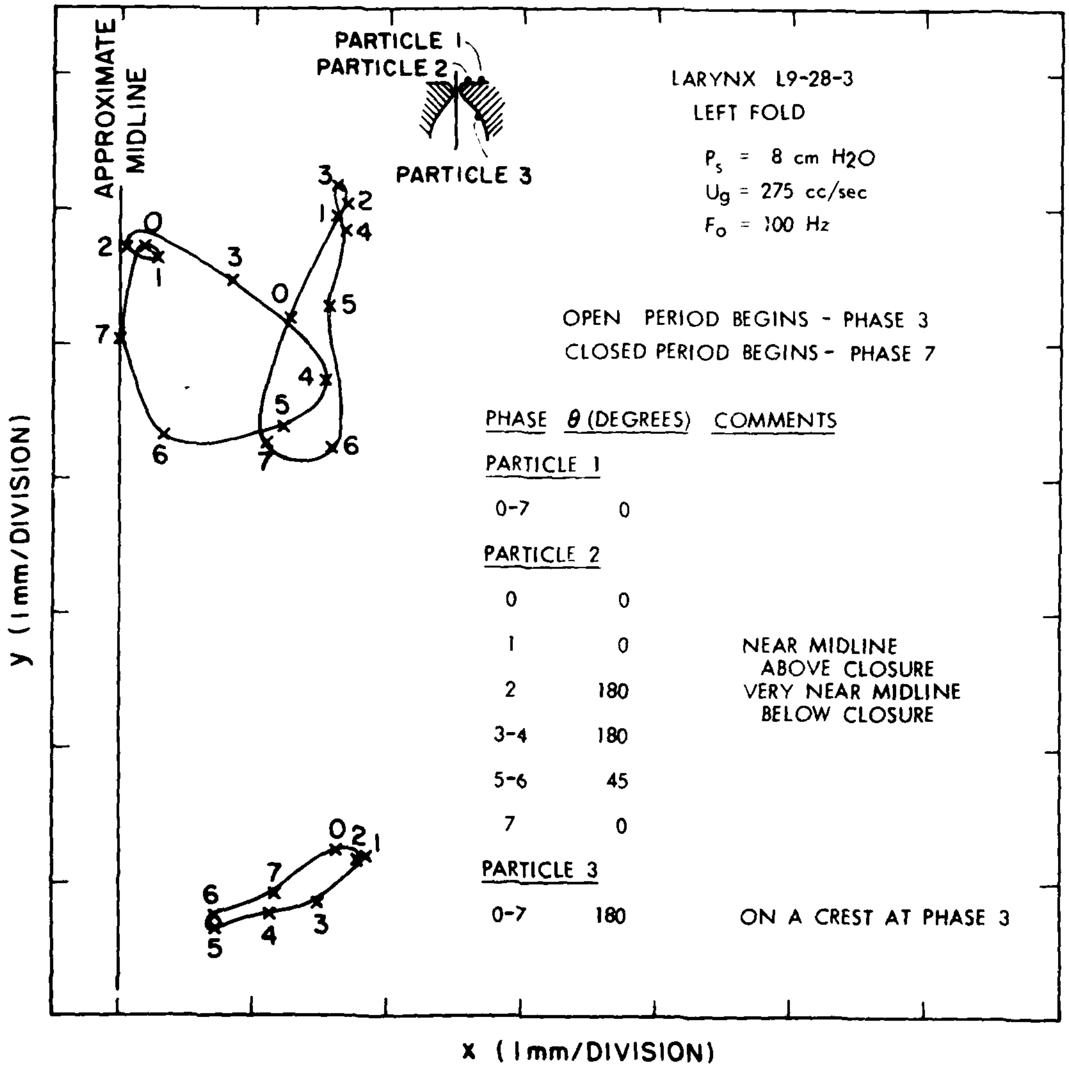


Figure 2. Frontal-plane trajectories of three particles during a single glottal cycle. Measurements were made at eighth cycle increments, numbered 0 through 7. The inset to the right of the trajectories contains notes about the measurements, including the angle, θ, of the tabletop for which each measurement was made. The schematic sketch at the top of the inset indicates the particle locations with respect to the margin of the vocal fold.

particles crossed, so that the particles were nearly vertically aligned during one measurement and horizontally aligned during another. Thus, the vibrations were complex. Some aspects of the trajectories and of vibrations in general were consistent with the notion of a displacement wave, progressing up the medial surface at a velocity of about 1m/sec, and then progressing laterally on the superior surface at .3-.5m/sec. The supraglottal wave was easily observed, as with normal human larynges, and its velocity was measured directly. Glottal closure also exhibited wavelike properties. Tissues at the lower edge of closure were peeled apart, while tissues above the point of closure were still coming together. The depth of closure was often almost negligible immediately before the glottis opened. The middle particle in Figure 2 appeared to be on the superior part of the vocal folds for part of the cycle, and was below the point of closure for part of the closed phase. Thus, it is evident that the vibrations are complex and cannot be well modeled, in detail, as simple translations of a small number of lumped-parameter masses.

Although some aspects of the vibration patterns seemed best describable by surface waves along the cover of the vocal folds, vibrations of the edge also appeared to be describable as string vibrations (that is, whole-body translation and torsional flexure). There may have been components of both types of vibrations. This interpretation is interesting, because interactions between the two types of vibration as a function of variations in control parameters may help to explain fine control over voice quality variations.

Detailed shapes of the vocal folds during the eight phase increments in Figure 2 were estimated and are shown in Figure 3. A two-mass model approximation could be superimposed on these shapes if vertical movements of the masses were allowed. Given this approximation, the aerodynamic theory of Ishizaka and Matsudaira (1972) was capable of reconciling average subglottal pressure with average flow rate. It was also shown, as expected, that the aerodynamic model provided for the efficient transfer of energy from the aerodynamic system to the mechanical system (Stevens, 1977), given the nature of vertical phase differences. The mechanical parts of the two-mass model did not well account for these data, however. Thus, to the extent it could be tested, the aerodynamic aspect of the two-mass model seemed accurate, but the mechanical part of the model seemed inadequate.

A change in particle trajectories was observed as the tissues desiccated and vibrations eventually ceased. These and other measurements suggested that particle trajectories could be considered as oscillations around an unstable equilibrium position. This result implies that small-signal modeling techniques, such as those of Ishizaka and Matsudaira (1972), which account for voice onset by finding unstable solutions to linear equations, are justified.

Excised larynges were able to produce nearly normal vibrations even when the vocalis muscle on one or both sides was completely removed. However, these preparations did not seem capable of falsetto vibrations. Wave motions with velocity similar to that of the normal case were still seen to propagate upward on the medial wall. Particle trajectories were somewhat similar to the normal case, although they differed in some details. These observations should be especially useful for testing models that account for the layered structure of the vocal folds.

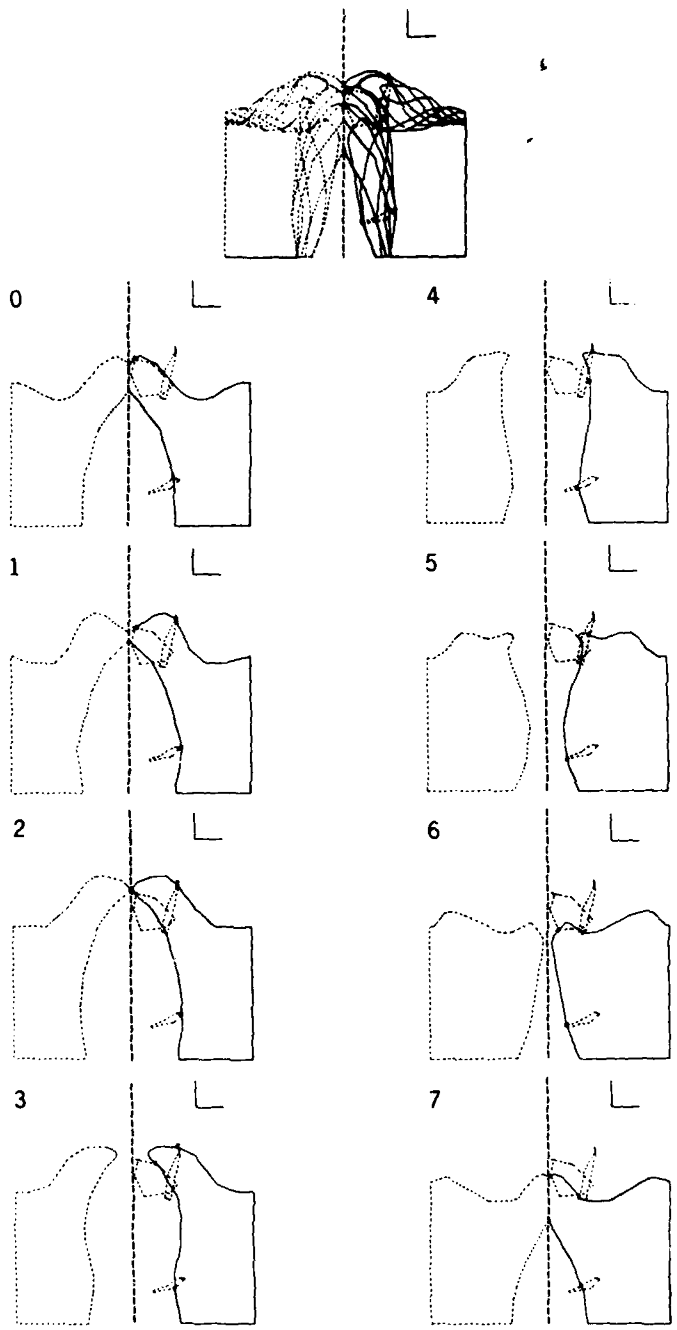


Figure 3. Sketches of vocal-fold shape during a vibratory cycle. These shapes were estimated on the basis of the data shown in Figure 2, which is superimposed in each panel. Bilaterally-symmetric shapes are shown for display purposes, although measurements were actually made on only one side. The corner in the upper right of each panel indicates 1 mm scales. Individual shapes at eighth cycle increments are shown at the lower part of the figure. The top panel shows all of them superimposed.

The experiments described above illustrate the potential value of developing a model specifically for excised larynges, as a step in developing a model for the in vivo case. An advantage to modeling the excised preparation explicitly is not only its versatility, as illustrated by the experiments with excised vocalis muscles, but also the fact that measurements of mechanical properties can be made on the same preparation on which the vibration patterns are measured.

Optical techniques for measuring frontal plane vibration patterns, such as those used by Baer, are limited because they are time consuming and because only vibrations of the vocal fold surfaces can be measured. Radiographic techniques may provide a solution to the problem of measuring vocal fold shapes throughout a cycle. There have been some radiographic studies of vocal fold vibrations in vivo. Sovak, Courtois, Haas, and Smith (1971) described a high-speed radiographic study capable of resolving the details of a glottal cycle. Hollień, Coleman, and Moore (1968) developed the technique of stroboscopic laminagraphy, in which an x-ray source is pulsed stroboscopically during a laminagraphic procedure. For steady phonation, images of a frontal section could thus be obtained at successive phases within a cycle. The usefulness of these studies was limited by the poor quality of the images obtained. Furthermore, they may be no longer practical, in view of modern concerns about radiographic dosage, especially to the thyroid gland. However, such techniques could be applied safely and more effectively to the study of excised or animal larynges. A promising improvement on these techniques was recently described by Saito (1977) and Saito, Fukuda, Ono, and Isogai (1978). Small lead pellets were affixed to the vocal fold surfaces and also implanted within the vocal folds, so that both internal and external vibrations could be monitored. Stroboscopic radiography, synchronized to the voice, was then used to track the movements of these particles throughout cycles of vibration. Such measurements might be made even more effectively with a computer-controlled x-ray microbeam system (Fujimura, Kiritani, & Ishida, 1973; Kiritani, 1977), if its detector output were stroboscopically sampled or its source stroboscopically pulsed, because of the improved spatial resolution of this device. Conceivably, radiopaque medium could be introduced through the circulatory system, as a further improvement of this technique.

II. MEASUREMENTS IN VIVO: RESPONSES TO INDIVIDUAL CONTROL VARIABLES

There are many parameters controlling phonation in the normal human larynx. Control is exerted most directly through the effects of the intrinsic muscles on laryngeal configuration and through transglottal pressure. Forces exerted by the extrinsic laryngeal muscles and other extrinsic structures also have an effect. Acoustic load can modify the patterns of airflow through the glottis and probably the mechanical vibrations as well. There are probably other effects, such as contol of vascular and mucous supply, which are less well understood. During voluntary control of phonation, variations in several of these parameters are intercorrelated (see, for example, Atkinson, 1978). Although such variables as the levels of electromyographic activity in individual muscles and subglottal pressure can be correlated with corresponding changes in fundamental frequency or other aspects of phonatory performance, correlation does not guarantee causality, because of the intercorrelations among control variables. Therefore, it has been difficult to isolate

the detailed phonatory response to any one of them. Nevertheless, these detailed effects must be known in order to determine the relevance of data from excised larynx and animal experiments, to adequately test detailed phonatory models, and, in general, to fully understand phonatory function.

One method for isolating the effects of a given parameter is to externally apply involuntary perturbations and observe the phonatory response while other parameters remain constant. This technique has been most successfully used for examining the effects of changes in subglottal pressure on fundamental frequency. Several experiments have been reported in which subglottal pressure is increased by a sudden push on the chest or abdomen of a phonating subject, and both subglottal pressure and fundamental frequency are monitored during an interval for which no muscular response is assumed to occur (for example, van den Berg, 1957; Isshiki, 1959; Ladefoged, 1963; Öhman & Lindqvist, 1966; Fromkin & Ohala, 1968). This experiment was recently replicated by Baer (1979), who also monitored the electromyographic activity of laryngeal muscles to ensure the absence of a response. Transglottal pressure can also be varied supraglottally, through modulation of intraoral pressure (Lieberman, Knudson, & Mead, 1969; Hixon, Klatt, & Mead; 1971; Rothenberg & Mahshie, 1977). When pressure modulations are oscillatory, at frequencies of about 6-10Hz, continuous muscular compensation does not seem to occur, although EMG evidence to support this claim has not been published.

Although results of these induced-pressure-change experiments differ in some details, their consensus indicates that fundamental frequency varies with transglottal pressure at rates of about 3-5Hzcm H_2O within the speech range, with higher rates at higher fundamental frequencies or in falsetto register. These results, as well as correlation between fundamental frequency and subglottal pressure during voluntary control (Atkinson, 1978), suggest that the phonatory response to pressure change is fast, perhaps within the interval of one or two glottal periods.

The effects of involuntary perturbations in acoustic load on fundamental frequency have also been investigated through systematic variation in the length of a tube that artificially extends the vocal tract (Ishizaka, Matsudaira, & Takashima, 1968; Ishizaka & Flanagan, 1972). Changes in fundamental frequency of as much as 20Hz were obtained by varying the length of the tube. However, it was not determined in these experiments whether there was any compensatory laryngeal response. It is easily shown that such artificially increased acoustic loads can have an effect on phonation. If one phonates an ascending scale into an artificially extended vocal tract (such as a mailing tube), the voice will typically break or switch to falsetto when the fundamental frequency nears the first resonance frequency of the tract. A lower order manifestation of this phenomenon might account for the intrinsic pitch of vowels (Peterson & Barney, 1952). In any case, such experiments could be repeated more carefully to further constrain the performance of phonatory models.

The logical counterpart to these studies for quantifying the effects of individual muscles on phonatory performance would probably require electrical stimulation of the muscles. There are no accounts of any such studies on normal human subjects, and it is unclear whether stimulation experiments are possible in practice. However, an alternative method, which isolates the

effects of single-motor-unit contractions, has recently been used by Baer (1978) for investigating the effects of individual muscles on fundamental frequency. Rather than analyzing gross aspects of fundamental frequency control, this method relates very small changes in fundamental frequency (namely, pitch perturbations) to very small changes in muscle tension, which can be related to single-motor-unit activity. Statistical independence between motor-unit inputs can then be exploited to uncorrelate the muscles, and examine their individual causal effects on fundamental frequency.

This method extends the use of an averaging technique that was first developed for studying properties of single motor units in skeletal muscles (Milner-Brown, Stein, & Yemm, 1973). Single-motor-unit action potentials (see Harris, 1981) must be identified in an electromyographic recording while the muscle sustains a contraction. A simplified muscle model, which is approximately valid at low to moderate levels of contraction, is assumed. This model is shown in Figure 4. Its inputs are the action potential trains from individual motoneurons. Each of these can be considered a random point process, and they are statistically independent across units. Each motor-unit action potential triggers a mechanical twitch--a positive pulse of tension whose detailed characteristics vary across motor units. At least some of these units fire at low enough rates so that adjacent twitches do not overlap. The output tension of the whole muscle is the summation of its constituent motor unit outputs. Although many of the motor unit outputs are trains of pulses, they sum to an approximately constant, though noisy, value because they are statistically independent. The relative amplitude of this noise depends on the number of motor units and their firing rates.

Given the model in Figure 4, the contribution of a single motor unit to the output tension (its contraction properties) can be estimated if its input action potentials can be identified and if these inputs are isolated by intervals great enough to ensure against overlap of adjacent contractions. Samples of the output tension waveform following the inputs are aligned and averaged. The output of the isolated motor units is always the same within these intervals, while the outputs of all other motor units are random and thus average to a constant value.

To apply this technique to investigation of fundamental frequency control, we note that motor-unit firings are statistically independent across muscles as well as within a muscle. We then hypothesize that muscle-tension variability contributes to the fundamental frequency perturbations that can be measured when a normal phonating subject attempts to sustain a steady tone. The resulting model for pitch perturbations is then indicated in Figure 5. Laryngeal muscles produce roughly constant output tensions that are noisy because of single-unit effects. The noise components across muscles are uncorrelated. The complex effect of muscle forces on the vocal folds, which we have lumped under the term "vocal fold tension," is also roughly constant, but noisy. Output fundamental frequency then depends on this tension and other independent inputs such as subglottal pressure and, perhaps, mucosity and other random effects. All the detailed inputs to this model are thus statistically independent. According to the model, then, fundamental frequency as a function of time can be treated as an output and be averaged just as muscle tension in earlier studies to estimate the effects of single-motor-unit contractions in that muscle. The effects of other muscles and other inputs average to a constant value.

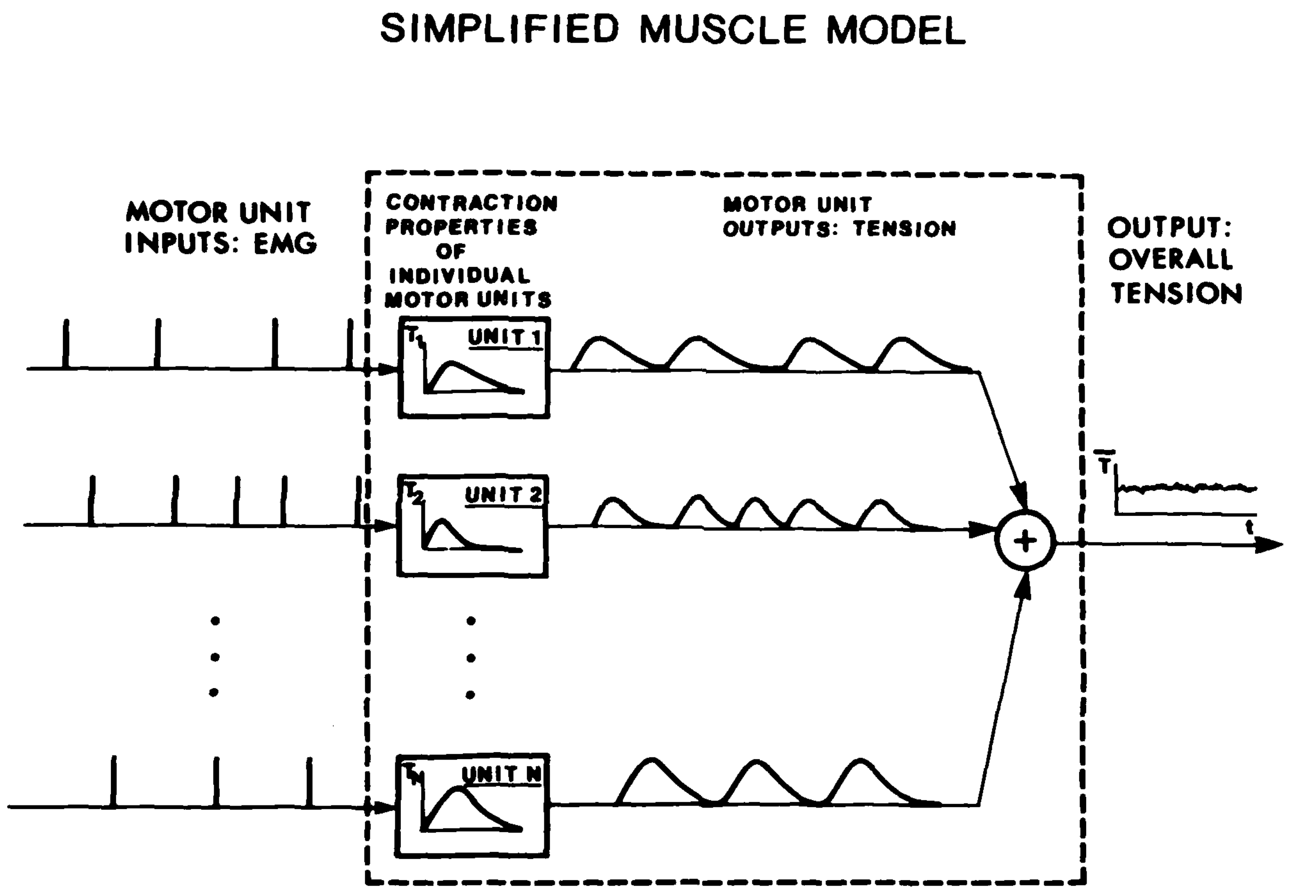


Figure 4. Simplified model of a muscle during a sustained contraction.

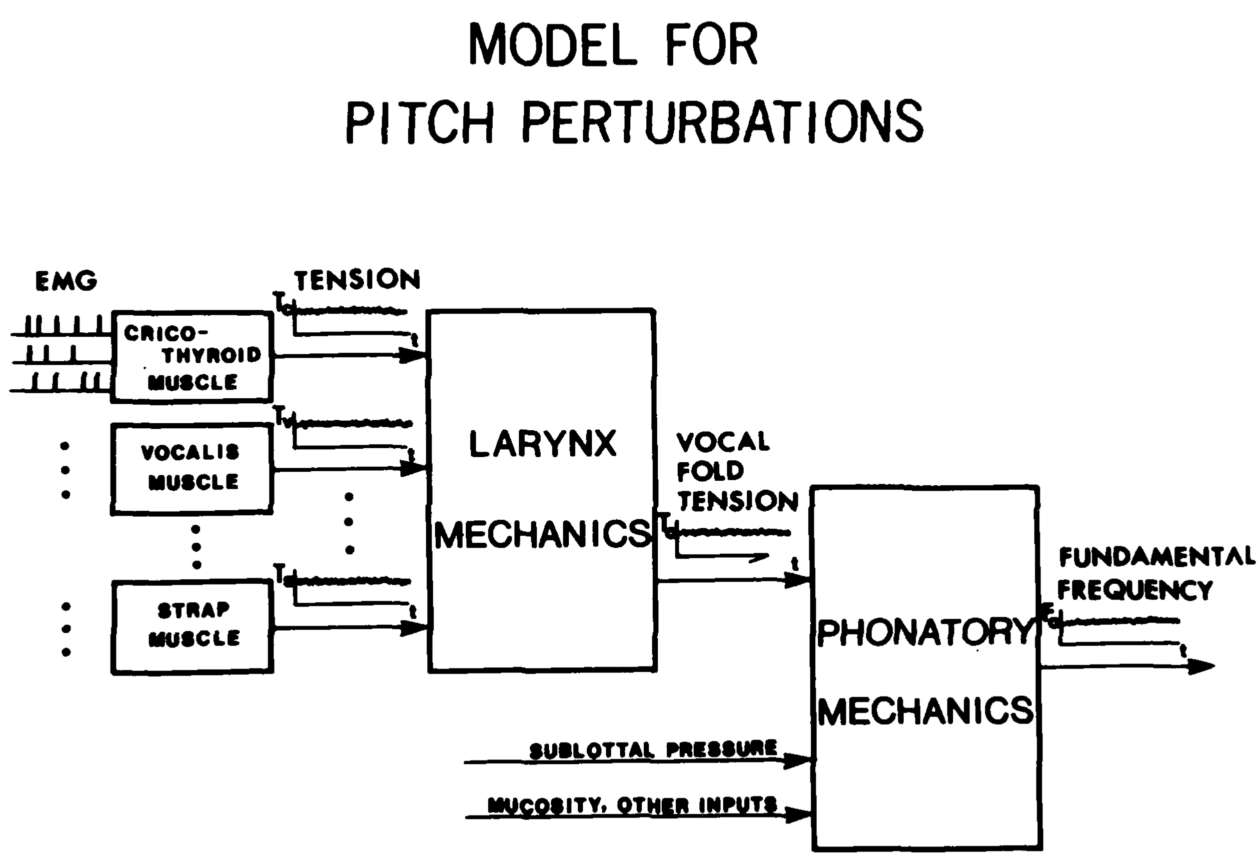


Figure 5. Model for pitch perturbations during production of a steady tone.

To obtain data for such a study, a subject is asked to sustain a steady tone for several breaths. Electromyographic (EMG) activity, obtained through hooked wire electrodes from a laryngeal muscle under study, and the voice signal obtained through a standard microphone are recorded and input to a digital computer. After instantaneous fundamental frequency as a function of time is derived, this waveform is offset by approximately its average value and amplified to exaggerate the perturbations. Isolated single-motor-unit firings are identified in the EMG waveform. Then, samples of the EMG waveform and the F_0 perturbation waveform are aligned around the single firings and averaged. The sample window extends from 100ms before to 300ms after these firings.

Figure 6 shows a 1.5s sample of data when the muscle under study was the cricothyroid, whose function as a vocal-fold tenser and hence as a pitch raiser is well known. Fundamental frequency was about 100Hz, which is in the lower part of the subject's range, in order to keep the number of recruited units and their firing rates low. As this figure shows, fundamental frequency was estimated to 1 Hz resolution. Although cycle-to-cycle variations rarely exceed 1Hz, perturbations over larger time intervals were about 4Hz wide. Two firings have been isolated in this record, and the corresponding sample intervals are indicated by horizontal lines.

Figure 7 shows the results of the averaging calculation for this experiment after 19 suitable firings were identified. The upper panel shows the averaged EMG signal, which exhibits a pulse only at the lineup point, as expected. The lower panel shows the average F_0 perturbation. This signal is approximately at baseline both to the left of the lineup point and to the far right of the window. However, there is a positive pulse beginning immediately after the lineup point. This pulse reaches its peak amplitude of 1Hz at a latency of about 70-80ms. The pulse appears to indicate that the single-motor-unit contraction caused, on the average, a 1Hz increase in fundamental frequency.

A similar calculation was performed for one of the strap muscles, an extrinsic laryngeal muscle whose possible function in lowering F_0 has been a source of some controversy. When fundamental frequency was in the middle of the subject's range, no systematic effect was found. Results when the fundamental frequency was low are shown in Figure 8. Although these data are somewhat noisier than those in Figure 7, they appear to exhibit a negative pulse in the interval immediately after the lineup point. Thus, the strap muscle is shown to have a causal effect in lowering fundamental frequency from an already low level.

The confirmation of a muscular contribution to F_0 perturbations is itself interesting, since perturbations have been used as an indicator of vocal pathology. These results show that care must be taken when interpreting patterns of perturbation. More relevant to this discussion, however, is the fact that we can show the response to a short duration pulse of tension in a single muscle, and that these data can thus be used to constrain the performance of laryngeal models. It was noted that the average pitch perturbation for the cricothyroid muscle begins immediately after the lineup point. This shows that the phonatory response must begin within one glottal cycle. The latency of the peak of the response, 70-80ms, includes contribu-

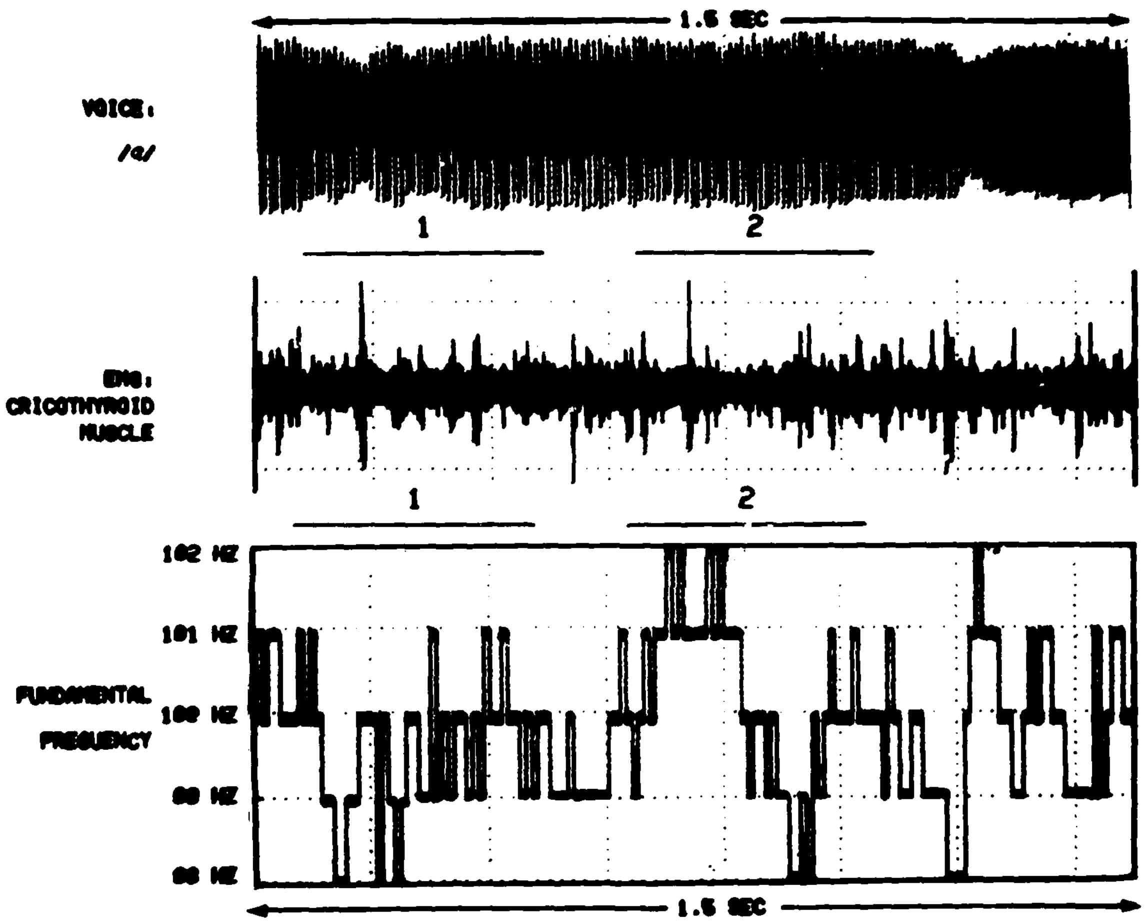


Figure 6. Short segment of data during production of a steady tone at about 100 Hz. Top: voice waveform; Middle: EMG activity of the cricothyroid muscle; Bottom: "instantaneous fundamental frequency" extracted from the voice waveform. Two sets of horizontal lines indicate intervals from 100 ms before to 300 ms after single-motor-unit firings in the cricothyroid muscle.

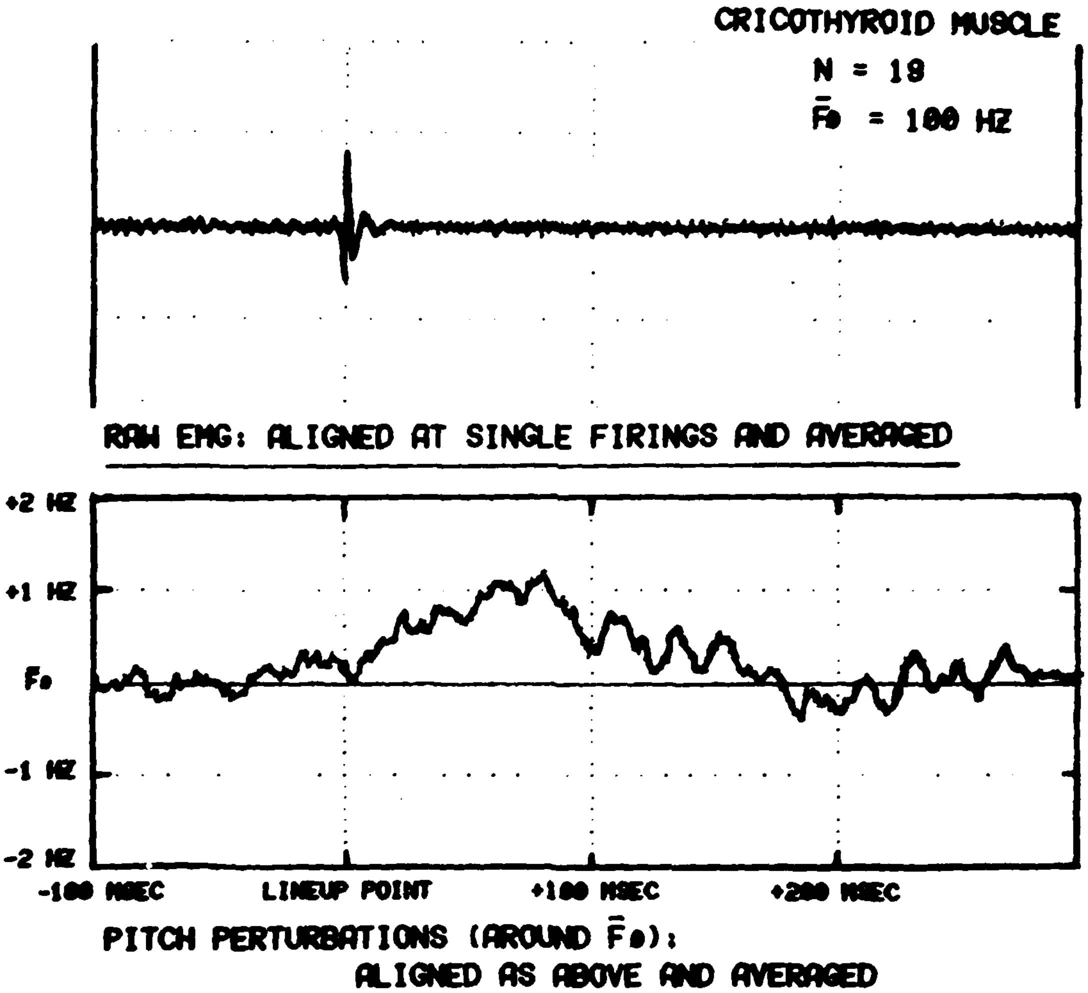


Figure 7. Ensemble-average waveforms of EMG activity from the cricothyroid muscle and corresponding instantaneous fundamental frequency. All waveforms have been aligned at the time of a single-motor-unit firing for purposes of averaging.

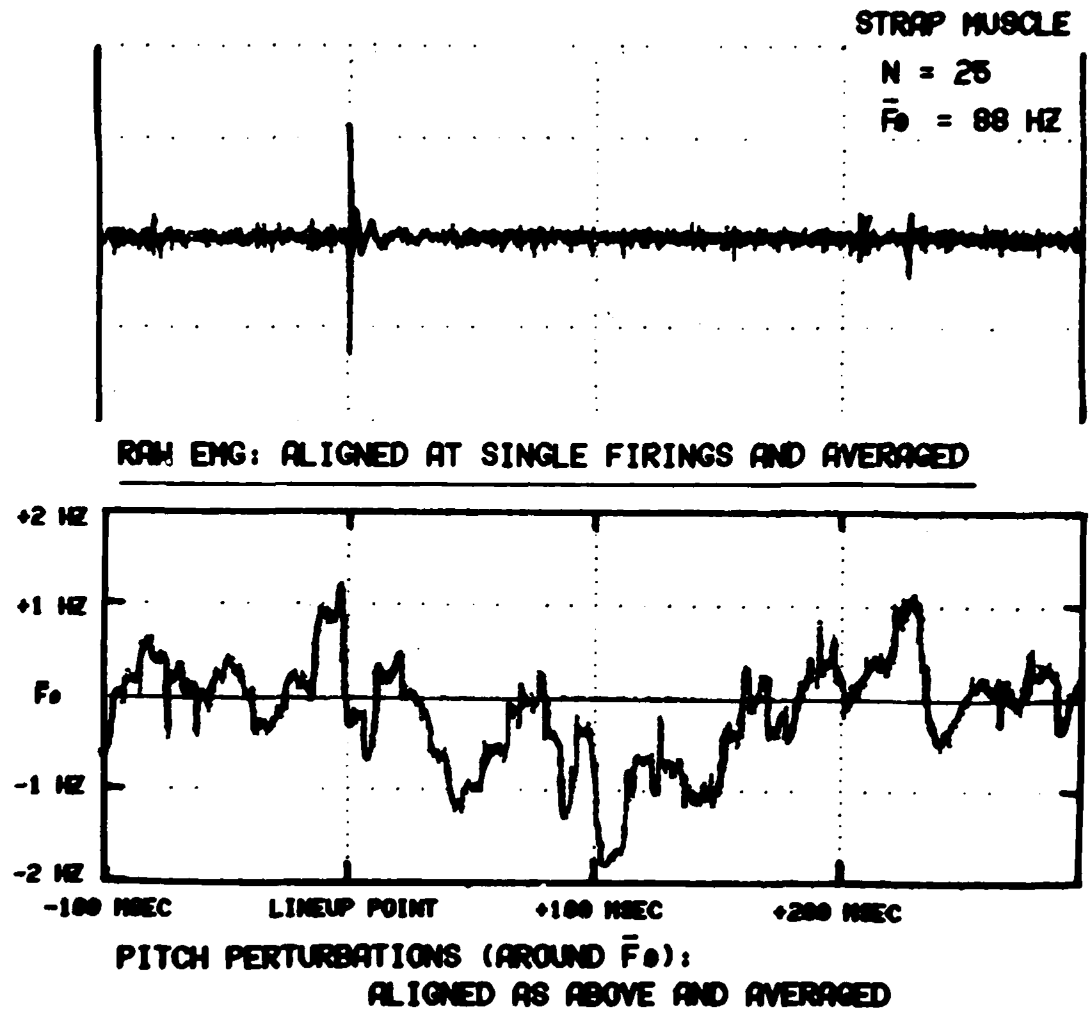


Figure 8. Ensemble-average waveforms of EMG activity from an unspecified strap muscle and corresponding instantaneous fundamental frequency. All waveforms have been aligned at the time of a single-motor-unit firing for purposes of averaging.

tions due to muscle contraction time, mechanical response latency in the larynx, and latency of phonatory response. Since both the latency and the amplitude of the mechanical motor-unit contractions can be estimated in animal experiments, these data might be further applied to the detailed testing of models of laryngeal performance, especially in comparison with data reported by Hirano (1975) relating changes in shape and mechanical properties of vocal folds to stimulation of various muscles. These data might also shed some further light on the pattern of motor control. For example, the relatively large amplitude of the F_0 perturbation pulse in Figure 7 relative to the overall perturbation in Figure 6 suggests that very few motor units were firing at rates low enough to show the effects of individual twitches. However, it is unclear how many other units may have been in tetanus. Perhaps the greatest value of the single-unit technique will be in elucidating the phonatory function of muscles such as the vocalis, whose gross patterns of activity are so intercorrelated with those of other muscles during ongoing regulation of phonation that their detailed effects have remained obscure.

In considering the function of individual control parameters in this section, we have only discussed measurements of their effects on fundamental frequency. The reason for this is that, with few exceptions, these are the only measurements that have been made. Fundamental frequency by itself, however, is evidently not a very complete descriptor of phonatory activity. As fundamental frequency is varied, attributes of the vocal source waveform that contribute to intensity and voice quality also vary. It is important to determine how these parameters covary when changes are produced by different control mechanisms, and, for purposes of assessing vocal pathology, how these relationships change in different pathological states.

Techniques to be discussed in today's session can be used to measure some of these different parameters of phonatory performance, such as amplitude of the glottal pulse and open quotient. When these parameters are measured cycle-to-cycle, the same techniques described in the section for studying fundamental frequency control can be utilized to assess the effects of different control parameters. These data, together with such anatomical and physical studies as those reported by Hirano (1975), are needed to improve our understanding of the phonatory mechanism and constrain the performance of mechanistic models. Thus, these studies should be pursued. Furthermore, if it were possible, it would be even more useful to study not only changes in vibratory performance characteristics as a function of these control parameters, but also intermediate variables such as the positions of the laryngeal structures and their mechanical properties. However, these experiments must await the development of techniques for measuring these parameters.

Finally, further insights are needed into the detailed conditions necessary for initiating and sustaining phonation, as well as for regulating ongoing phonation. An example of how such studies might be performed in vivo is by using involuntary perturbations of subglottal pressure. For example, a subject might be asked to assume a configuration appropriate for voicing but to maintain subglottal pressure at a level below the threshold for voice onset. Transglottal pressure might then be suddenly increased, say using a chest push procedure, to a level for which phonatory vibrations are initiated, while laryngeal configuration remains constant. Conditions for voice onset could then be determined, in terms of the level of subglottal pressure, as a

function of variations in the configuration. With negative transglottal pressure perturbations, conditions for voice offset could also be studied.

REFERENCES

Atkinson, J. E. Correlation analysis of the physiological factors controlling fundamental voice frequency. Journal of the Acoustical Society of America, 1978, 63, 211-222.

Baer, T. Investigation of phonation using excised larynxes. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1975.

Baer, T. Effect of single-motor-unit firings on fundamental frequency of phonation. Journal of the Acoustical Society of America, 1978, 64, S90. (Abstract)

Baer, T. Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. Journal of the Acoustical Society of America, 1979, 65, 1271-1275.

Fant, G. Acoustic theory of speech production. 's-Gravenhage: Mouton, 1960.

Flanagan, J. L., & Landgraf, L. Self-oscillatory source for vocal-tract synthesizers. IEEE Transactions on Audio and Electoacoustics, 1968, AU-16, 57-64.

Fromkin, V., & Ohala, J. Laryngeal control and a model of speech production. UCLA Working Papers in Phonetics, 1968, 10, 98-110.

Fujimura, O., Kiritani, S., & Ishida, H. Computer controlled radiography for observation of movements of articulatory and other human organs. Computers in Biology and Medicine, 1973, 3, 371-384.

Harris, K. S. Electromyography as a technique for laryngeal investigation. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, this volume.

Hirano, M. Morphological structure of the vocal cord as a vibrator and its variations. Folia Phoniatrica, 1974, 26, 89-94.

Hirano, M. Phonosurgery: Basic and clinical investigations. Otologia Fukuoka, 1975, 21, 239-440.

Hirano, M. Structure and vibratory behavior of the vocal folds. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977, 13-27.

Hixon, T. J., Klatt, D. H., & Mead, J. Influence of forced transglottal pressure changes on vocal fundamental frequency. Journal of the Acoustical Society of America, 1971, 49, 105. (Abstract)

Hollien, H., Coleman, R., & Moore, P. Stroboscopic laminagraphy of the larynx during phonation. Acta Oto-laryngologica, 1968, 65, 209-215.

Ishizaka, K., & Flanagan, J. L. Synthesis of voiced sound from a two-mass model of the vocal cords. Bell System Technical Journal, 1972, 51, 1233-1268.

Ishizaka, K., & Matsudaira, M. Fluid mechanical considerations of vocal cord vibration. SCRL Monograph (Speech Communication Research Laboratory, Santa Barbara), 1972, No. 8.

Ishizaka, K., Matsudaira, M., & Takashima, M. Variations of the vocal pitch with the driving point impedance. Journal of the Acoustical Society of Japan, 1968, 24, 313-314.

Isshiki, N. Regulatory mechanism of the pitch and volume of voice. Oto-Rhino-Laryngology (Kyoto), 1959, 52, 1065-1094.

Kiritani, S. Articulatory studies by the x-ray microbeam system. In

M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977, 171-190.

Ladefoged, P. Some physiological parameters in speech. Language and Speech, 1963, 6, 109-119.

Lieberman, P., Knudson, R., & Mead, J. Determination of the rate of change of fundamental frequency with subglottal air pressure during sustained phonation. Journal of the Acoustical Society of America, 1969, 45, 1537-1543.

Matsushita, H. Vocal cord vibration of excised larynges--A study with ultra-high speed cinematography. Otologia Fukuoka, 1969, 15, 127-142.

Milner-Brown, H. S., Stein, R. B., & Yemm, R. The contractile properties of human motor units during voluntary isometric contractions. Journal of Physiology, 1973, 228, 285-306.

Öhman, S., & Lindqvist, J. Analysis-by-synthesis of prosodic pitch contours. Quarterly Progress and Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), 1966, STL-QPSR 1/1966, 1-6.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 1952, 24, 175-184.

Rothenberg, M., & Mahshie, J. Induced transglottal pressure variations during voicing. Journal of the Acoustical Society of America, 1977, 62, S14. (Abstract)

Saito, S. Phonosurgery. Otologia Fukuoka, 1977, 23, 171-384.

Saito, S., Fukuda, H., Ono, H., & Isogai, Y. Observation of vocal fold vibration by x-ray stroboscopy. Journal of the Acoustical Society of America, 1978, 64, S90-91. (Abstract)

Smith, S. On artificial voice production. In A. Sovijarvi & P. Aalto (Eds.), Proceedings of the 4th International Congress of Phonetic Sciences. The Hague: Mouton, 1962, 96-110.

Sovak, M., Courtois, J., Haas, C., & Smith, S. Observations on the mechanism of phonation investigated by ultraspeed cinefluorography. Folia Phoniatrica, 1971, 23, 277-287.

Stevens, K. N. Physics of laryngeal behavior and larynx modes. Phonetica, 1977, 34, 264-279.

Titze, I. R. The human vocal cords: A mathematical model. Part 1. Phonetica, 1973, 28, 129-170.

Titze, I. R. The human vocal cords: A mathematical model. Part 2. Phonetica, 1974, 29, 1-21.

Titze, I. R., & Talkin, D. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. Journal of the Acoustical Society of America, 1979, 66, 60-74.

van den Berg, Jw. Sub-glottal pressure and vibrations of the vocal folds. Folia Phoniatrica, 1957, 9, 65-71.

van den Berg, Jw. Myoelastic-aerodynamic theory of voice production. Journal of Speech and Hearing Research, 1958, 1, 227-244.

van den Berg, Jw., & Tan, T. S. Results of experiments with human larynges. Practica Oto-Rhino-Laryngology, 1959, 21, 425-450.

van den Berg, Jw., Zantema, J. T., & Doornenbal, P. On the air resistance and the Bernoulli effect of the human larynx. Journal of the Acoustical Society of America, 1957, 29, 626-631.

PHONETIC PERCEPTION OF SINUSOIDAL SIGNALS: EFFECTS OF AMPLITUDE VARIATION*

Robert E. Remez,+ Philip E. Rubin, and Thomas D. Carrell++

Abstract. Naive subjects, when instructed to listen for a sentence, are capable of transcribing the phonetic message of acoustic signals consisting solely of time-varying sinusoids. These unnatural-sounding signals mimic the pattern of formant center-frequency and amplitude variation over the course of polysyllabic, semantically normal utterances. To what extent does amplitude variation over time contribute to intelligibility? Our present investigation tested the hypothesis that listeners derive some information about syllable patterns from amplitude variation alone, and may therefore use contextual constraints to deduce prosodically appropriate portions of the message in the tonal stimulus. Phonetic and syllabic intelligibility were compared in four conditions: (1) normal amplitude and frequency variation; (2) normal frequency variation with constant amplitude; (3) normal frequency variation with a misleading amplitude contour; and (4) normal amplitude variation with no frequency variation. These results are discussed in the framework of phonetic perception and in terms of current theories of the perception of fluent speech.

Talkers make sounds for listeners to hear. This truism has implicitly motivated many present explanations of speech perception. Essentially, these explanations have sought to enumerate the perceptually critical acoustic elements produced by talkers when generating phonetic sequences. Researchers have used the ability to synthesize speech to fashion acoustic signals containing only those acoustic components of natural utterances believed to be necessary for perception. In doing so, we have made highly refined and specific descriptions of the stimuli that elicit phonetic perception. In complementary research, studies of the auditory periphery, of the basilar membrane, cochlear nucleus and auditory projection have permitted us to learn how the critical acoustic elements survive auditory transmission. But,

---

*Paper presented at the 101st Meeting of the Acoustical Society of America, Ottowa, Ontario, Canada, May 22, 1981.
+Department of Psychology, Barnard College, Columbia University, New York, New York.
++Department of Psychology, Indiana University, Bloomington, Indiana.
Acknowledgement. For helping us conceptually, we thank Franklin Cooper, Alvin Liberman, David Pisoni, Brad Rakerd, and Michael Studdert-Kennedy. This research is supported by a grant from Sigma Xi to Robert E. Remez, Grant HD 01994 from the National Institute of Child Health and Human Development to Haskins Laboratories, and Grant MH 24027 from the National Institute of Mental Health to David B. Pisoni.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

regardless of the differences among the many approaches to studying phonetic perception, all approaches have assumed that the stimuli for phonetic perception consist necessarily of the kinds of sounds produced by a variably excitable, variably shapable tube-resonator--the vocal tract.[1]

A recent demonstration of ours questioned the assumption that the perceiver requires phonetic stimuli to comprise, however selectively, acoustic elements found in natural utterances (Remez, Rubin, Pisoni, & Carrell, 1981). In raising this question, our study also challenged the assumption that phonetic perception is based simply on a succession of discrete acoustic elements. In this study, we used a signal consisting of three time-varying sinusoids, each of which varied in a way that a formant peak might vary over the course of an utterance. Initially we fabricated the sinusoidal pattern by computing the resonant center-frequencies of a natural utterance, using Linear Predictive Coding (see Figure 1). The table of values produced through this analysis was used to set frequency and amplitude parameters of a sine-wave synthesizer. Figure 2 shows the differing short-time Fourier spectra of natural, synthetic (OVE and Haskins Pattern Playback), and sine-wave signals. Note the absence of a fundamental frequency, harmonic spectrum, and broadband formants in the sinewave signal. Lacking these acoustic attributes, the sinewave spectrum does not resemble the spectrum of a natural signal, in any literal sense. However, there _is_ energy, albeit infinitely narrowband, at the computed peaks throughout the duration of the pattern; and, the time-varying properties of the sinewave pattern, specifically the coherence of the _changes_ of the energy peaks over time, replicate the natural case.

The perceptual effects of sinewave stimuli were easy to predict. Because the short-time spectra of three-tone signals differ drastically from natural and even synthetic speech; because no talker is capable of producing three simultaneous "whistles" with these bandwidths, in this frequency range; and because the frequency and amplitude variation of the three tones is not synchronized, the perceiver should hear three independent streams, one for each sinusoid. The perceiver should hear no phonetic qualities.

However straightforward this prediction seems, there was a second, contrasting prediction. Suppose that the listener is able to disregard the short-time differences between sinusoidal signals and speech, and can attend, instead, to the overall pattern of change of the three tones. The pattern of change of the frequency peaks resembles the resonance changes produced by a vocal tract articulating speech. If the listener can apprehend this coherence in the time-varying properties of the nonspeech signal, then he should hear a phonetic message spoken by an impossible voice.

Given nonspeech stimuli whose time-varying properties are abstractly vocal, listeners perceived the signals in _both_ of the ways we predicted. Those listeners who were told nothing about the stimuli heard science fiction sounds, bad electronic music, sirens, computer bleeps and radio interference.[2] Those listeners who instead were instructed to transcribe a "strangely synthesized English sentence" did exactly that, for the most part--they identified the radically unnatural "voice" quality of the patterns, but they transcribed those patterns as they would have the original natural utterances upon which we based our sinewave stimuli.

SINEWAVE SYNTHESIS SIMULATION

OF A NATURALLY PRODUCED UTTERANCE

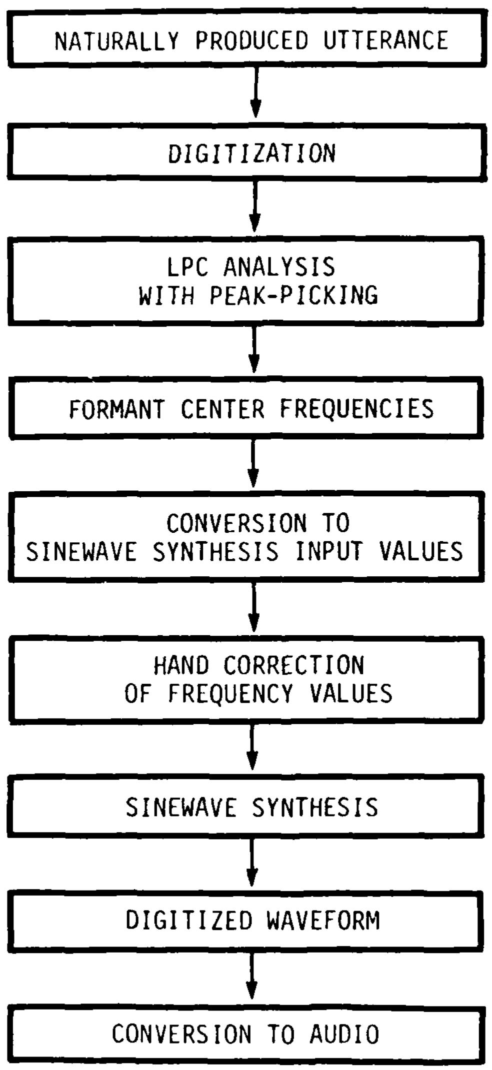


Figure 1. Sinewave stimuli are produced by imitating the time-varying properties of the center frequency and amplitude of the first three formants in a natural utterance.

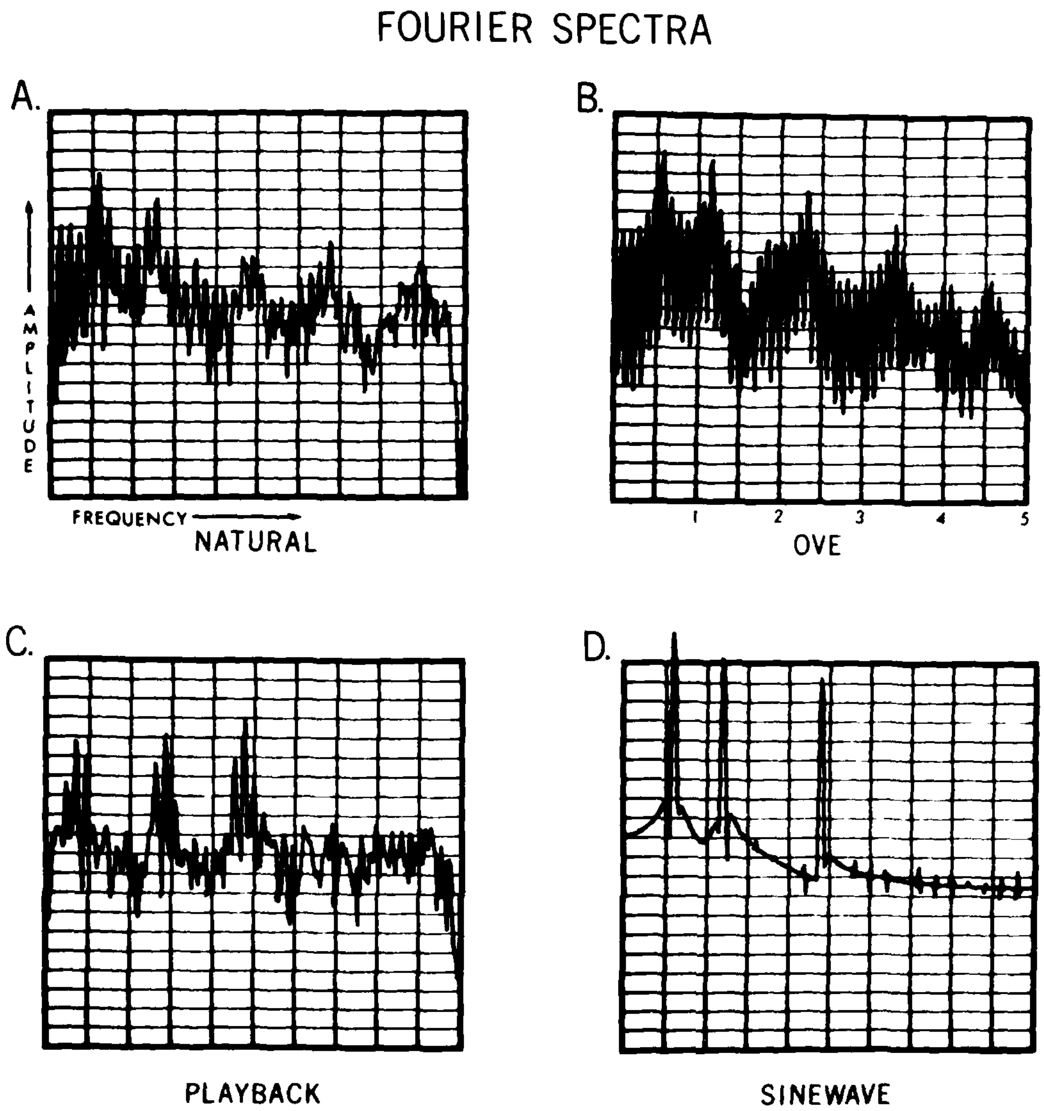


Figure 2. A comparison of the Fourier spectrum of four complex waveforms. (A) natural speech; (B) synthetic speech produced by the OVE synthesizer; (C) synthetic speech produced by the Haskins Labs Pattern Playback; (D) waveform consisting of three sinusoids.

This finding was novel in at least two ways. (1) It extended research on phonetic perception of sinusoidal signals to a high uncertainty judgment task, by offering unrestricted response alternatives. Previous tests of sinusoidal patterns had used forced-choice identification tasks with small response sets (Bailey, Summerfield, & Dorman, 1977; Best, Morrongiello, & Robson, 1981; Cutting, 1974; Fant, 1959; Grunke & Pisoni, 1979). Subjects' performance is obviously stabilized in such circumstances. However, we showed that the intelligibility of sinusoids does not depend on extensive training with simple, schematic stimuli, nor on test procedures that intrinsically promote consistent performance.

(2) More generally, the study indicated that speech perception is possible despite drastic departures from the short-time spectra of natural speech--despite absence of broadband formants, harmonic spectrum, and fundamental frequency--insofar as the time-varying properties of speech signals are preserved; and, insofar as the listener is able to attend to the coherent time-variation of the acoustic pattern. Both of these general qualifications must obtain for phonetic perception of sinusoids to occur, for the listeners who were not directed to expect speech for the most part did not spontaneously hear phonetic sequences in the tones.

The present investigation is directed toward questions that arose from our initial research with perception of sinusoidal replicas of fluent, semantically ordinary utterances. Primarily, we noted that the tonal patterns could well be considered an extreme case of defective acoustic-phonetic stimuli. If this description were apt, then the perceptual process could be described more conventionally, in quite different terms. Listeners might merely have memorized the tune of the tones without any phonetic recognition; and, after inferring a prosodic schema from the amplitude contour preserved in the tonal pattern, listeners would then have been free to guess (or, rather, to hypothesize) a likely phonetic sequence for the utterance using "top-down" finesse. A number of views of the perception of fluent speech include a prominent faculty for best-guessing lexical patterns from the prosodic structure when the phonetic stimulus is defective or ambiguous (e.g., Cutler & Foss, 1977; Huggins, 1978; Nakatani & Schaffer, 1978). Perhaps the listeners in our original study relied on such guesswork for transcribing the stimulus, and did not immediately perceive the message from phonetic structure preserved in the time-varying tonal pattern. In that case, very little phonetic perception would have occurred, and our theoretical claim would need to be moderated.

In the test we report here, each listener was presented with a sinusoidal pattern replicating the sentence "Where were you a year ago?" In response, the listener reported two things: (1) a transcription of the sentence; and (2) a count of the syllables in the sentence. If phonetic information is preserved in the coherence of the changing sinusoids, then transcription performance should be no poorer than syllable counting, which would presumably be based here on the linguistic structure of the message. If, on the contrary, only prosodic information in the form of amplitude variation is readily available to the listener, then syllable counting should be much more accurate than transcription of the message. In this latter condition, subjects would be likely to vary in the particular phonetic guesses they make given that an infinity of sentences may conform to the same prosodic pattern.

The present test also included a stimulus manipulation to evaluate more directly the difference between perceiving the phonetic structure and guessing about it based on amplitude information about prosody. Four conditions were used. In the first, listeners gave their two responses to a sinusoidal pattern that preserved both peak-frequency and peak-amplitude change of the first three formants of the original, natural utterance (see Figure 3). In the second condition, listeners heard a pattern that preserved the frequency variation of the first three formant center-frequencies at a constant level of energy throughout the utterance (see Figure 4). In the third condition, the sinusoidal pattern preserved the frequency pattern of the first three formants, but with a grossly misleading amplitude contour containing four segments of high energy and five segments of low energy, high and low differing by approximately 20dB (see Figure 5). The fourth condition employed a sinusoidal pattern with the original formant amplitude variation but with no frequency variation (see Figure 6). If the coarse amplitude structure of the stimuli provides reliable prosodic structure, and if subjects rely on this source of information about the message, then syllable counting should be accurate in conditions 1 and 4, and poorer in conditions 2 and 3. In addition, the accuracy of transcription should follow the accuracy of counting. If subjects perceive the phonetic sequence based on the time-varying properties of frequency variation, however, transcription and counting should be good in all conditions but the fourth, in which there is no frequency variation.

Our results are straightforward, as Figure 7 depicts. Transcription was good in conditions 1 ($n=14$), 2 ($n=13$) and 3 ($n=12$); there was no statistical effect of the amplitude manipulation in these conditions. This indicates that subjects were not hindered by defective coarse acoustic structure when fine acoustic structure was available for phonetic perception. (Condition 4 was not scored for transcription, for the obvious reason that there was nothing phonetic to transcribe.) In the syllable counting task, there was an enormous difference between condition 4 (no frequency variation, appropriate amplitude variation) and the other three conditions (appropriate frequency variation with either normal, flat, or misleading amplitude variation). A post hoc means test confirmed that this effect is highly significant (Scheffe, $p<.001$). Subjects were clearly unable to derive syllable information solely from amplitude variation in this case (cf. O'Malley & Peterson, 1966).

We conclude from these results that sinusoidal signals do not consist of veridical prosodic information and defective acoustic-phonetic information. Listeners lacked the ability to follow the syllable structure when only the amplitude variation of the original transcribable pattern was preserved, yet they were able to apprehend the phonetic detail even when the energy contour was grossly inappropriate to the segments within it. It seems that listeners who transcribed these sinusoidal replicas of speech must have relied on information about the phonetic sequence available in the frequency variation alone.

Overall, these studies of sinusoidal signals contribute new knowledge about phonetic perception that is perhaps counterintuitive. That is, phonetic perception can be elicited solely by a coherent pattern of acoustic variation comprising elements that cannot, in principle, be realized vocally. In order to detect this coherence despite unproducible short-time spectra, listeners

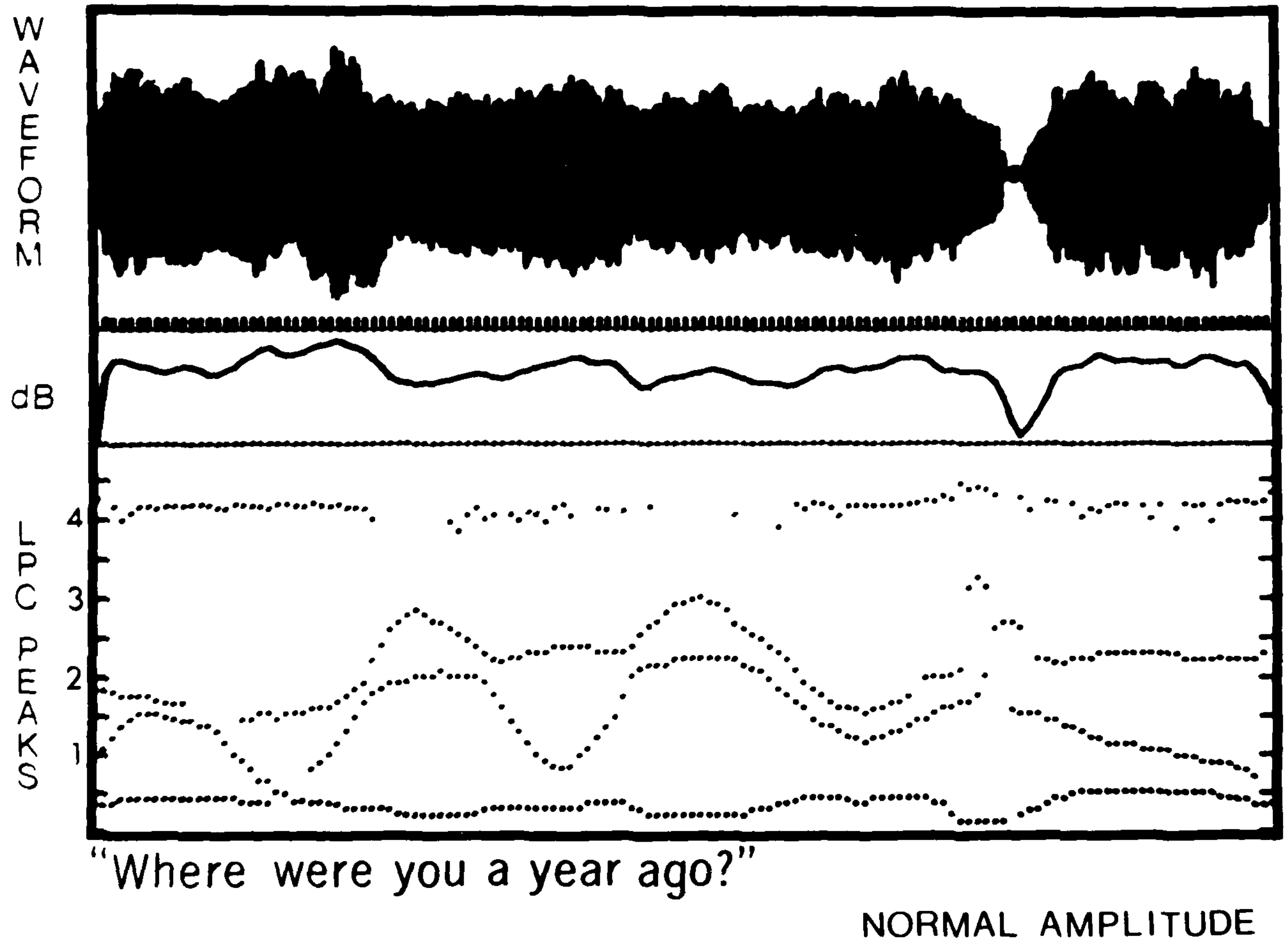


Figure 3. Display of waveform, energy and frequency change of three-tone replica of "Where were you a year ago?" Stimulus condition 1.

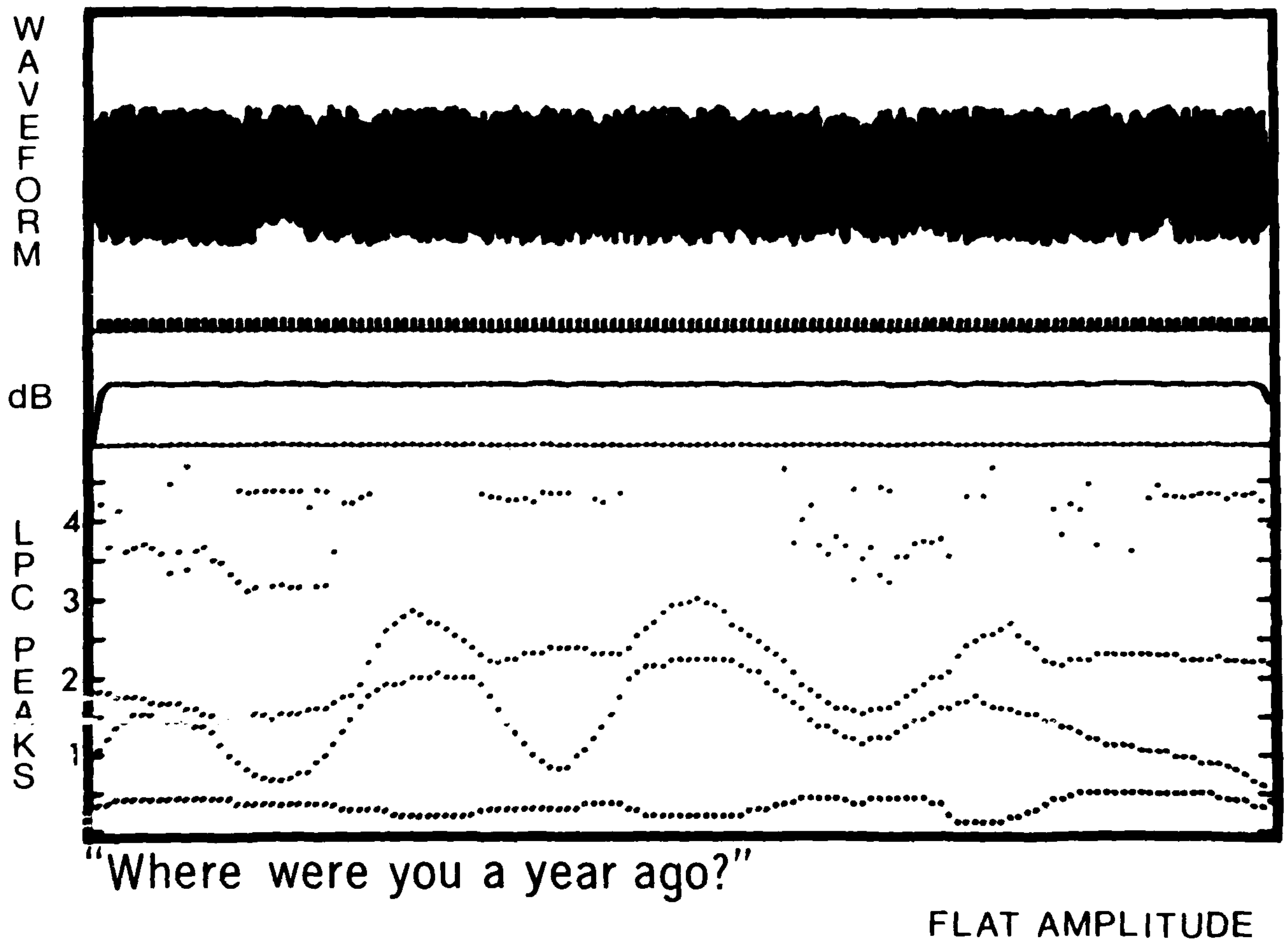


Figure 4. Stimulus condition 2: variation in the frequency of the three tones at a constant energy level.

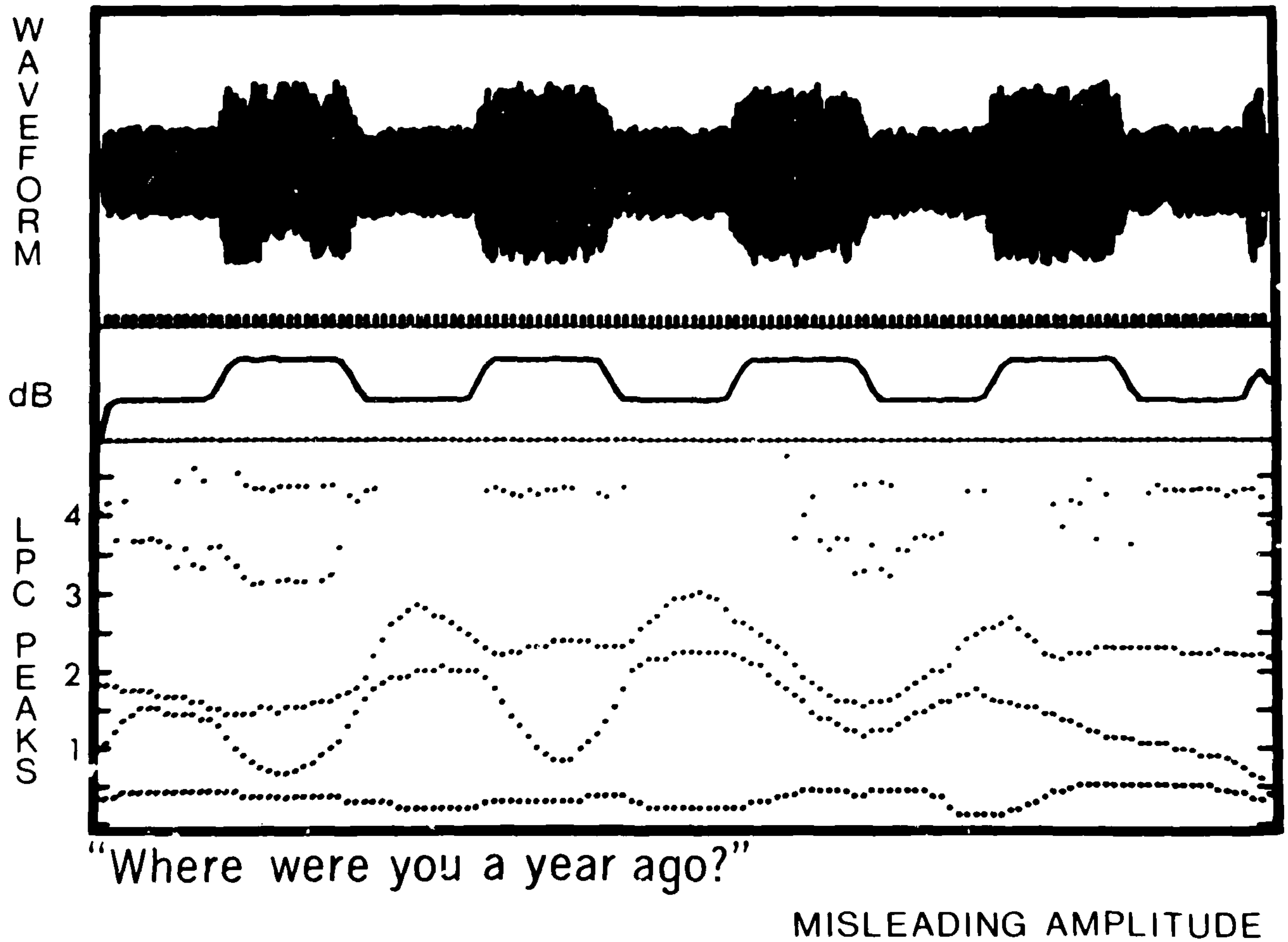


Figure 5. Stimulus condition 3: variation in the frequency of the three tones with a prosodically misleading amplitude pattern.

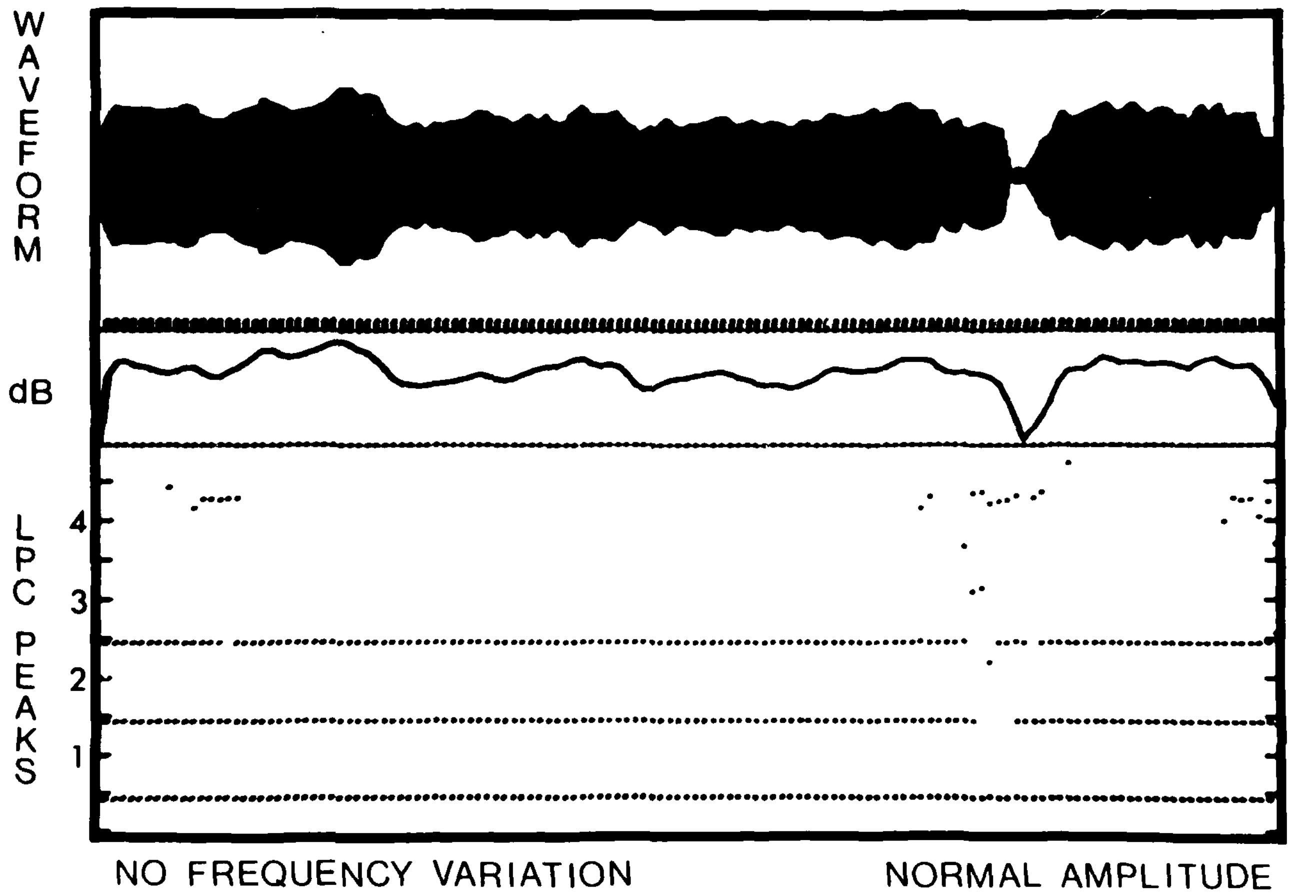


Figure 6. Stimulus condition 4: no frequency variation with the prosodically appropriate amplitude pattern.

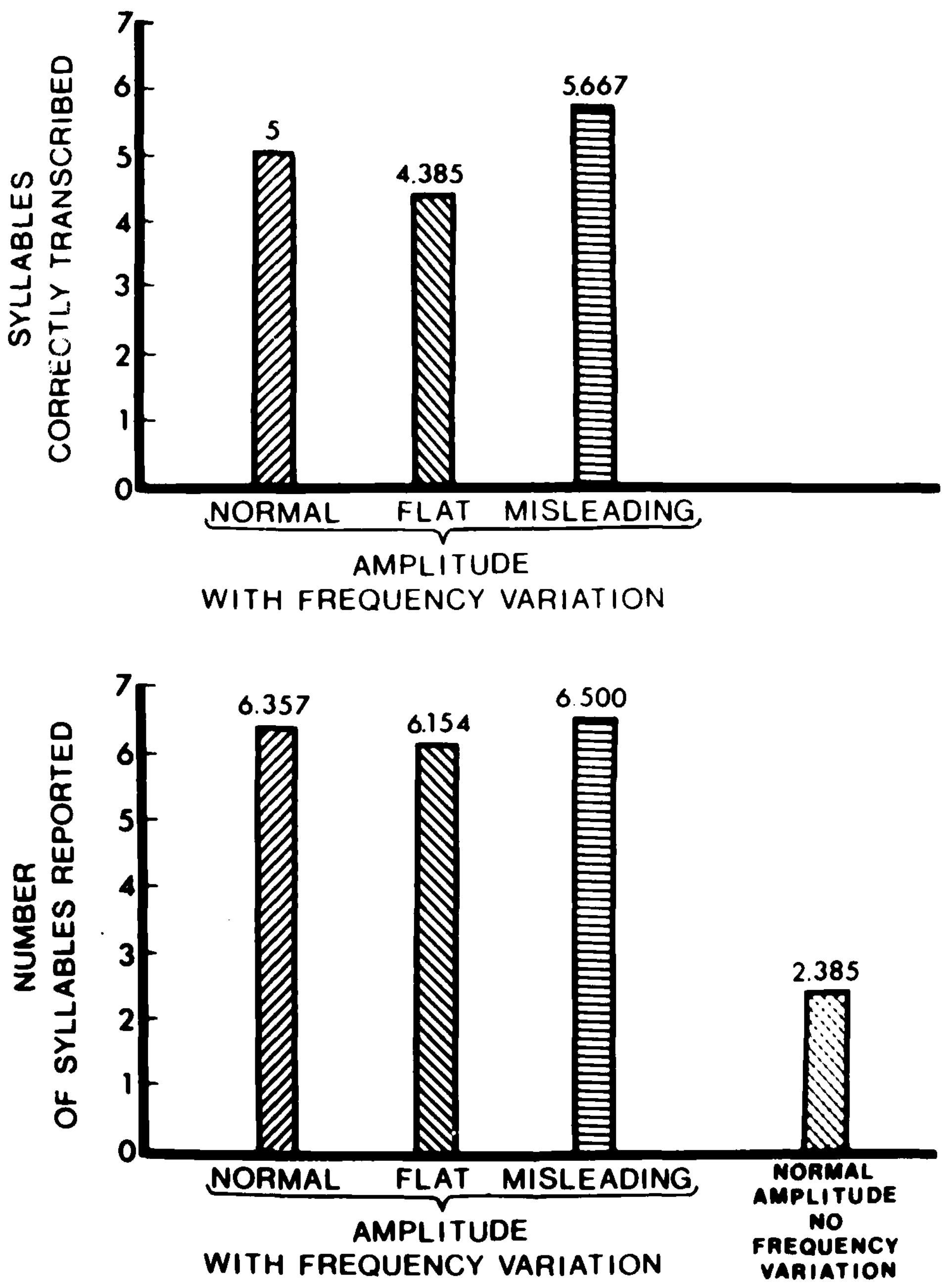


Figure 7. Top: group averages of transcription performance. Bottom: group averages of syllable counting.

must ultimately rely on even more abstract and more forgiving knowledge of vocal tracts than has been proposed by Liberman (1979). We venture to say that phonetic perception may actually be based on attention to the coherent patterns of change in acoustic energy rather than on attention to the particular qualities of the successive, discrete acoustic elements that compose the speech signal. To refine our speculation, we must extend this technique to a wider phonetic repertoire; to a more varied test of short-time spectral properties that permit the effect to occur; and to manipulations of the coherence of change directly.

REFERENCES

Bailey, P. J., Summerfield, A. Q., & Dorman, M. On the identification of sine-wave analogues of certain speech sounds. Haskins Laboratories Status Report on Speech Research, 1977, SR-51/52, 1-25.

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 1981, 29, 191-211.

Cutler, A., & Foss, D. J. On the role of sentence stress in sentence processing. Language and Speech, 1977, 20, 1-10.

Cutting, J. E. Two left-hemisphere mechanisms in speech perception. Perception & Psychophysics, 1974, 16, 601-612.

Fant, G. Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics, 1959, 15, 3-108.

Grunke, M. E., & Pisoni, D. P. Perceptual learning of mirror-image acoustic patterns. In E. Fischer-Jørgenson, J. Rischel, & N. Thorsen (Eds.), Proceedings of the Ninth International Congress of Phonetic Sciences (Vol. 2). Copenhagen: Institute of Phonetics, 1979, 461-467.

Huggins, A. W. F. Speech timing and intelligibility. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978, 279-298.

Liberman, A. M. How abstract must a motor theory of speech perception be? Revue de Phonétique Appliquée, 1979, 49/50, 41-58.

Nakatani, L. H., & Schaffer, J. A. Hearing "words" without words: Prosodic cues for word perception. Journal of the Acoustical Society of America, 1978, 63, 234-245.

O'Malley, M. H., & Peterson, G. E. An experimental method for prosodic analysis. Phonetica, 1966, 15, 1-13.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. Speech perception without traditional speech cues. Science, 1981, 212, 947-950.

FOOTNOTES

[1]To our knowledge, no one claims that the properties of a talker's utterances necessary to perception are supplied in the auditory channel, though such a view cannot be excluded a priori.

[2]A very small number of listeners did recognize some phonetic properties of the stimuli.

MEMORY FOR ITEM ORDER AND PHONETIC RECODING IN THE BEGINNING READER*

Robert B. Katz,+ Donald Shankweiler,+ and Isabelle Y. Liberman+

Abstract. A defect in immediate memory for item order is often attributed to poor beginning readers. We have supposed that this problem may be a manifestation of an underlying deficiency in the use of phonetic codes. Accordingly, we expected good and poor readers to differ in their ability to order stimuli that can be easily recoded as words and stored in phonetic form, but not in their ability to order nonlinguistic stimuli that do not lend themselves to phonetic recoding in short-term memory. The purpose of the present study was to test this hypothesis by examining the ability of good and poor readers to reconstruct the order of sets of briefly presented stimuli that varied in the extent to which they could be distinctively recoded into phonetic form: pictures of common objects versus nonrepresentational, "doodle" drawings. As expected, an interaction between reading ability and type of stimulus item was found, demonstrating the material-specific nature of poor readers' ordering difficulties. These findings support the hypothesis that a function of the phonetic representation is to aid in retention of order information, and that poor readers' ordering difficulties are related to their deficient use of phonetic codes.

Certain commonly occurring memory problems of poor beginning readers have been regarded as manifestations of an underlying deficiency in the use of phonetic codes. Several studies have shown that children who are poor readers tend to make ineffective use of phonetic coding in short-term recall of linguistic material (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). However, special difficulties with recall and recognition arise only when the stimulus items are words or other items that can readily be labeled linguistically and retained phonetically in working memory (Holmes & McKeever, 1979; Vellutino, Pruzek, Steger, & Meshoulam, 1973; Vellutino, Steger, & Kandel, 1972). When the stimuli do not lend themselves to phonetic coding, the performances of good and poor readers cannot be distinguished. For example, we (Liberman, Mann, Shankweiler, & Werfelman, Note 1) tested recognition memory with two sets of stimuli that could not be easily labeled:

---

*To appear in Journal of Experimental Child Psychology.
+Also University of Connecticut.
Acknowledgment. This investigation was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. We are grateful to the principal and teachers of the Parker Memorial School, Tolland, Connecticut, for allowing us to work with the second-grade classes. We are also grateful to the children who participated and their parents for their cooperation. Special thanks are due to Leonard Katz for statistical advice.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

unfamiliar faces and abstract, nonrepresentational line drawings (Kimura, 1963). It was found that good and poor readers were indistinguishable on memory for both faces and nonsense drawings.

The question we ask here is whether children's memory for the order of occurrence of stimulus items would also vary with their phonetic recodability. Repeatedly, the literature has suggested that poor readers have difficulty in retaining the order of items in tests of serial recall (Bakker, 1972; Benton, 1975; Corkin, 1974). There are indications, as we noted, that the poor readers' deficits in item recall may be a manifestation of their deficient ability to use phonetic codes. We should now ask whether the deficits they might have in remembering the order of stimuli would also vary with the phonetic recodability of the items. This is what we would expect in light of suggestions that one function of phonetic memory codes is to preserve item order (Baddeley, 1978; Crowder, 1978). Consequently, we would suppose that the poor reader's difficulty in retaining order information is material-specific and not a global memory deficit for item order.

To pursue this question experimentally, we needed to discover how poor readers would fare with order memory for nonlinguistic material. While it is true that some studies (Corkin, 1974; Noelker & Schumsky, 1973; Stanley, Kaplan, & Poole, 1975) have reported inferior performance by poor readers in ordering nonlinguistic stimuli, the interpretation of the findings in each case is open to some question either because the items used were such as to be readily labeled or were presented for long exposure times. In either instance, even though the stimuli presented were nonlinguistic, the effect of the procedure might be to accentuate the differences in performance between the reader groups by encouraging linguistic recoding on the part of the good readers who habitually recode phonetically. Moreover, good and poor readers have been found to be equivalent in ordering other nonlinguistic items, such as photographed faces (Holmes & McKeever, 1979). At all events, there has been no direct test of the hypothesis that the poor readers' problem with order memory may be linked to a deficiency in the use of phonetic codes. The present experiment was designed to provide direct evidence for such a link. By controlling for the ease with which linguistic labels can be given to test items, we expected to find that differences in the performances of good and poor readers would depend on the phonetic recodability of the stimulus material.

The experiment compared good and poor readers' memory for order for two sets of controlled stimuli: a set consisting of items that are easily labeled --line drawings of common objects, and a set containing items presumed to be very difficult to label--Kimura's (1963) nonsense drawings. The latter were chosen for use in this study because good and poor readers performed equally well with these stimuli in the test of recognition memory to which we referred earlier (Liberman et al., Note 1).

In the present procedure, a linear array of five figures is tachistoscopically presented, after which copies of the five figures are presented on cards, one figure per card, in random order. Subjects are asked to rearrange the cards, reconstructing the order in the previous display. Since poor readers tend not to make full use of phonetic coding in working memory, we expected them to be less accurate than good readers in ordering the

phonetically recodable pictures of common objects, but not to differ from the good readers in ordering the nonrecodable, doodle drawings. Thus we expected an interaction between reading ability and stimulus type, attributable to differences in the degree of reliance on phonetic recoding.

METHOD

Subjects

Subjects were selected from four second-grade classes in the Tolland, Connecticut public school system. Candidates for the poor reader group were selected for screening if they were so designated by their teachers or if they scored at the 40th percentile or lower on both word recognition subtests of the *Comprehensive Test of Basic Skills* (CTBS) (1974), which had been administered in the seventh month of the first grade. Candidates for the good reader group either received a superior evaluation from the teachers or ranked at or above the 80th percentile on both CTBS subtests.

Subjects selected for screening were administered the *Slosson Intelligence Test* (Slosson, 1963) and the word identification and the word attack subtests of the *Woodcock Reading Mastery Tests* (Woodcock, 1973) in the fifth and sixth months of the school year. The final good reader group consisted of those subjects who attained a combined raw score of at least 115 on the two Woodcock subtests, while the poor reader group included subjects with a combined score of less than 85. Subjects with extreme IQ scores (below 90 or above 135) were ineligible for further testing. In addition, one poor reader had to be dropped because of prolonged absence and ensuing scheduling difficulties. By these criteria, 21 good readers (10 females, 11 males) and 21 poor readers (7 females, 14 males) were selected. The good readers had a mean age of 95.1 months compared to the poor readers' mean age of 97.2 months, $t(40) = 1.7$; $p = .10$. The good readers had a mean IQ of 115.3 while the poor readers had a mean IQ of 107.4, $t(40) = 2.7$; $p = .012$. The mean combined raw score on the Woodcock was 134.6 for the good readers (range: 118 to 153) and 53.0 for the poor readers (range: 22 to 77).

Stimuli and Apparatus

Two sets of 50 drawings comprised the stimuli of this study. The first set consisted of the 50 nonsense drawings of Kimura (1963), which we designate "phonetically unrecodable" because they are difficult to label distinctively. The second set, which we call "phonetically recodable," included 50 line drawings of common objects. The latter had been shown in earlier pilot studies to be easily recognized by second graders, each drawing typically eliciting a single response which was a monosyllabic word. Each stimulus condition required 20 test trials. Each trial consisted of a tachistoscopic presentation of a different horizontal array of five stimuli mounted on 2 x 2 inch slides. To generate the required 20 arrays for each condition, 10 arrays were selected by random drawing without replacement from the set of 50 stimuli for that condition. Then 10 more arrays were generated by a second drawing for each stimulus condition. One set of three stimuli not used in the test trials was prepared to be used as practice trials. A sample array for each stimulus condition is displayed in Figure 1.

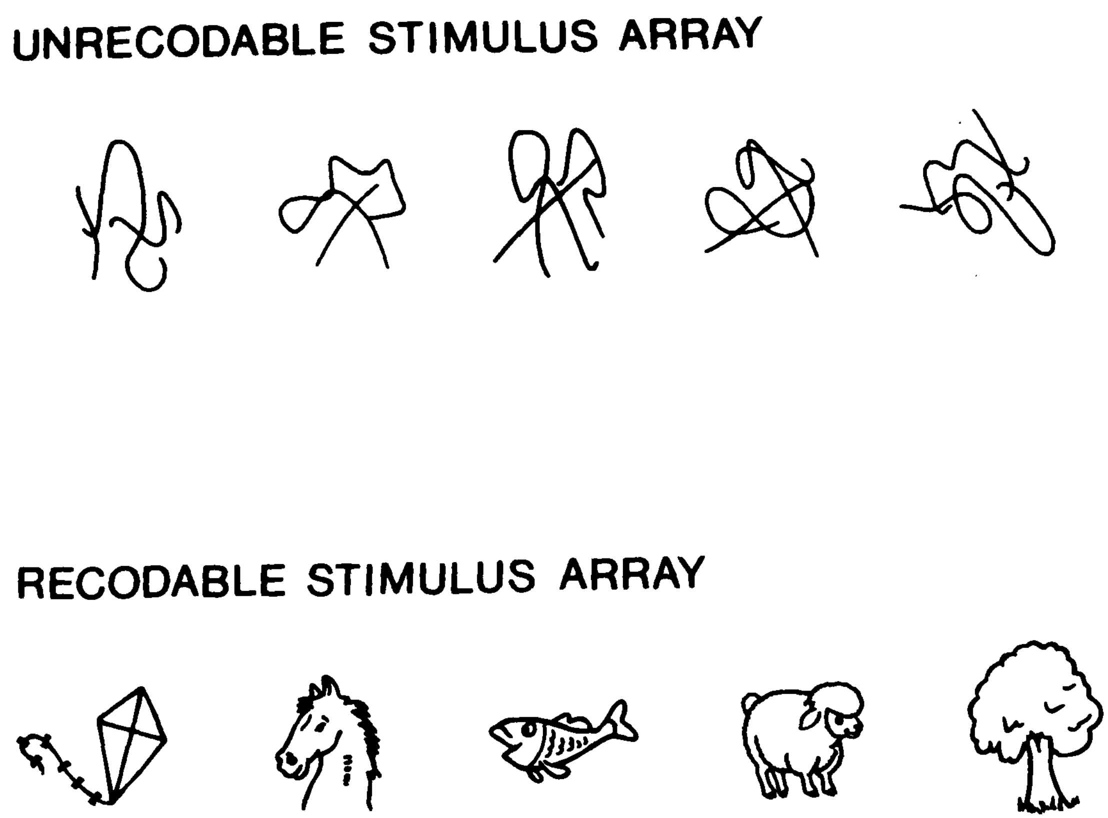


Figure 1. The upper portion of the figure gives a sample stimulus array consisting of five nonrepresentational line drawings (adapted from Kimura, 1963) for which ready verbal labels are not available. The lower portion gives a sample array for the comparison condition in which the items are easily named common objects (adapted from Makar, 1969).

The stimuli were projected onto a white screen for 4.0 sec using a carousel projector equipped with a tachistoscope attachment and a decade interval timer. The projected array was viewed from about 55 inches and extended a horizontal distance of about 15 inches (15.5 degrees). Each stimulus array subtended a visual angle of 1.5 to 2.3 degrees horizontally and 1.0 to 2.3 degrees vertically. A permanent focal point of reflective tape was attached to the left of the projected stimulus array.

For the ordering task, each stimulus item was individually reproduced on a laminated, white 3 x 5 inch card.

Procedure

Subjects were tested individually in two separate sessions, one session for each stimulus condition. The two sessions were conducted on separate days. To guard against transfer of a phonetic recoding strategy from one session to the next, the initial session was always devoted to the phonetically unrecodable condition.

Subjects were informed that they would see five figures on the screen for a brief period of time after which they would have to rearrange copies of the figures on the table in the same order. To provide some control for the direction of eye movements, subjects were instructed to fixate on the taped focal point before each trial. Immediately after each tachistoscopic presentation, a sheet of cardboard on the table was removed to reveal the five stimulus cards appropriate to that trial, arranged in random order. The same order was used for corresponding trials across the two conditions. No time limit was placed on the subject's performance. In both conditions, a rest period of approximately 2 min followed the tenth trial.

In each condition, a practice trial of three stimuli was presented before the 20 test trials. If the subject failed to order the stimuli correctly on the practice trial, the trial was repeated once. In any case, the practice set was always reviewed with the subject to insure that the task was understood.

RESULTS

The number of stimuli correctly ordered by each subject for each condition was tallied for all serial positions. To be considered correct, a stimulus item had to be placed in the serial position that corresponded to its original position on the slide. Figure 2 shows the mean number correct at each serial position for each group of subjects. It is clear from inspection of the group data depicted in the figure that both good and poor readers performed better with the easily recodable stimuli. This result obtained for every individual subject as well. It is also apparent from the figure that the average difference between the good and poor readers' performances was small in the unrecodable condition, compared to the corresponding difference in the recodable condition. In the phonetically unrecodable condition, poor readers averaged 5.6 stimuli correct per serial position, compared to the good readers' 6.7, while in the phonetically recodable condition, poor readers averaged 11.1 correct compared to the good readers' 14.1.

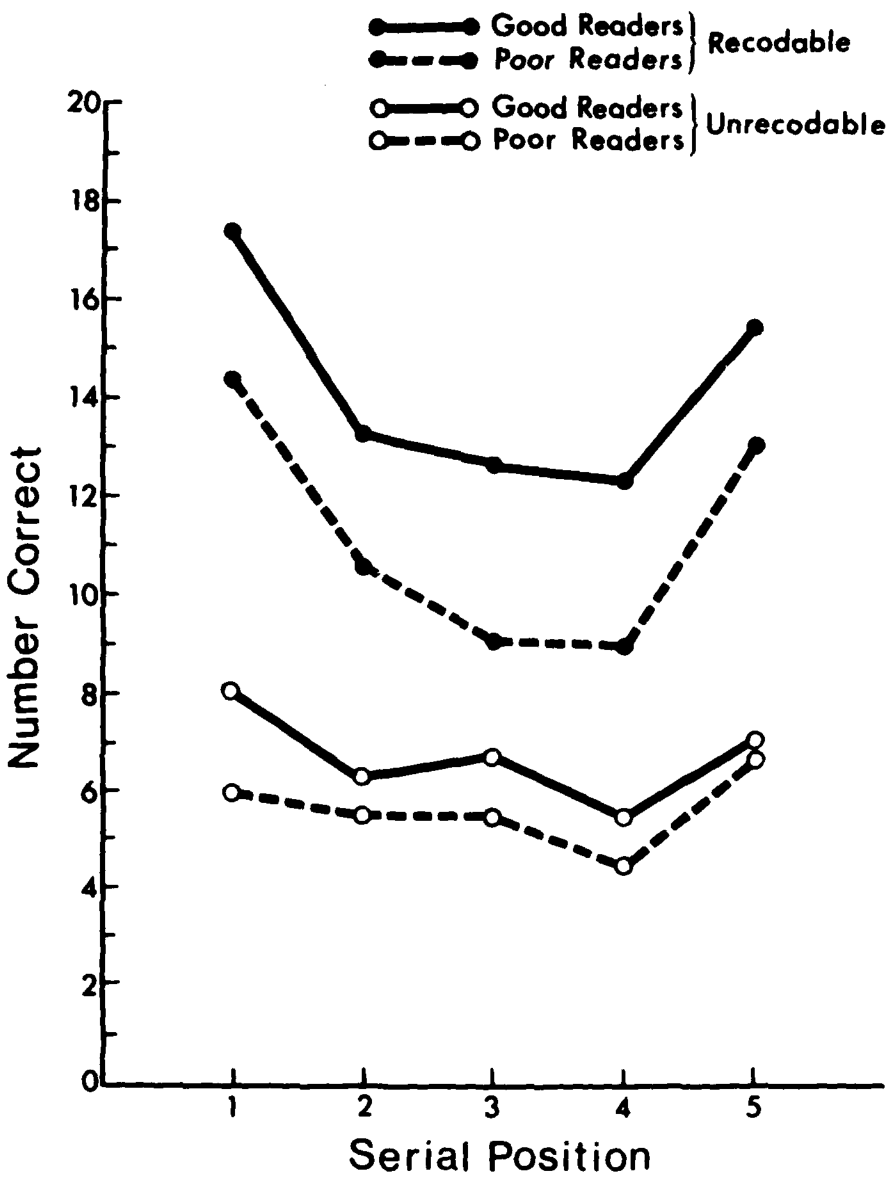


Figure 2. The mean number of items correctly ordered is plotted by serial position in the stimulus array. Separate curves are shown for each group on each task.

The data were subjected to an analysis of variance with one between-groups measure (reading ability) and two within-groups measures (stimulus recodability and serial position). All three main effects were highly significant: reading ability, $F(1,40) = 22.4$, $p < .001$; stimulus recodability, $F(1,40) = 236.1$, $p < .001$; and serial position, $F(4,160) = 30.9$, $p < .001$. The variation in shape of the serial position curves with a change in stimulus recodability is indicated by the interaction between stimulus recodability and serial position, $F(4,160) = 11.2$, $p < .001$. Of special interest was the interaction between reading ability and stimulus recodability, $F(1,40) = 5.1$, $p = .03$, confirming that the difference in performance between good and poor readers varies with recodability of the stimuli. A more fine-grained analysis of the interaction using a protected t-test (Cohen & Cohen, 1975) demonstrated that the mean performances of good and poor readers in the unrecodable condition were not significantly different, $t(40) = 0.8$, $p = .58$. In contrast, a significant difference was found in the recodable condition, $t(40) = 2.3$, $p = .028$.

An analysis of covariance using IQ as the covariate indicated that IQ was not significantly correlated with performance on the experimental task. The significant interaction between reading ability and stimulus recodability with IQ controlled, $F(1,39) = 5.0$, $p = .032$, argues against attributing the obtained differences in performance to differences in intelligence between the good and poor readers of our sample.

However, the rather low level of performance of all the subjects on the unrecodable condition raises the question as to whether the interactions obtained may have been falsely inflated by a floor effect. A floor effect would be expected to constrain the variance of the scores on the more difficult task. Therefore, the standard error of the means of the scores at each serial position on the two tasks was examined for indications of heterogeneity. It was found that the standard error for the scores on the unrecodable condition ranged from 0.31 to 0.66, whereas for the recodable condition, the standard error ranged from 0.55 to 0.78. Thus, since the ranges of these measures of variability differed for the two tasks, it is possible that the reading ability-by-stimulus recodability interaction that had been obtained might indeed have been falsely inflated.

This finding prompted us to do a further analysis, this time on the final ten trials alone. This portion of the data was selected on the assumption that previous practice may have brought the performances sufficiently above chance on the unrecodable condition to remove any constraining effects on the variance. As can be seen in Table 1, the number of correct placements (averaged over serial position) did increase for both groups in the unrecodable condition. Moreover, the heterogeneity of variance is completely eliminated in these final ten trials. For these trials, the standard error of the mean for the scores on the unrecodable condition ranged from 0.22 to 0.50 (poor readers: 0.30 to 0.45; good readers: 0.22 to 0.50); for the recodable condition, the standard error ranged from 0.26 to 0.49 (poor readers: 0.31 to 0.47; good readers: 0.26 to 0.49). Since heterogeneity of variance is clearly not a problem here, we can be more confident that any possible interactions involving the recodability factor would not be artifactual.

Table 1

Number of Correct Placements in Each Condition (Averaged over Serial Position) for the Initial Ten Trials and the Final Ten Trials

Stimulus Condition	Trials 1-10		Trials 11-20	
	Poor Readers	Good Readers	Poor Readers	Good Readers
Unrecodable	2.7	2.9	2.9	3.9
Recodable	5.8	6.9	5.3	7.2

Performances of good and poor readers on the final ten trials were then subjected to the same analysis as had been carried out on the full data set. An analysis of variance was computed with one between-groups measure (reading ability) and two within-groups measures (stimulus recodability and serial position). This analysis again revealed significant main effects of reading ability, $F(1,40) = 28.7$, $p < .001$, stimulus recodability, $F(1,40) = 200.8$, $p < .001$, and serial position, $F(4,160) = 24.8$, $p < .001$. In addition, the interaction between stimulus recodability and serial position was again obtained, $F(4,160) = 3.8$, $p = .006$. Finally, and most importantly, the interaction between reading ability and stimulus recodability was once more significant, $F(1,40) = 5.5$, $p = .025$. Moreover, with IQ controlled in an analysis of covariance, the latter interaction remained significant, $F(1,39) = 5.3$, $p = .027$. Post hoc analyses using protected t-tests (Cohen & Cohen, 1975) once more demonstrated that the performances of good and poor readers in the unrecodable condition were not significantly different, $t(40) = 1.6$, $p = .12$, whereas a significant difference was found in the recodable condition, $t(40) = 3.0$, $p = .004$.

DISCUSSION

We have raised the possibility that the problems in memory for order often imputed to poor readers may be a consequence of deficient use of phonetic memory codes. This possibility was explored by requiring subjects to reconstruct from memory the order of one set of stimuli consisting of drawings of easily named, common objects and another set consisting of nonrepresentational, doodle drawings that do not readily lend themselves to linguistic labeling. The results confirmed our expectations: the performances of good and poor readers did not differ significantly when the task required them to order stimuli that are difficult to label, but good readers were significantly better than poor readers in ordering stimuli that

are amenable to labeling. Since items that are labeled by words would be available to a phonetically-based working memory, the results are consistent with earlier indications of good readers' superior ability to make use of phonetic coding in working memory (Liberman et al., 1977; Mann et al., 1980; Mark, Shankweiler, Liberman, & Fowler, 1977; Shankweiler et al., 1979).

The fact that all subjects performed the ordering of the nonsense designs much less accurately than the ordering of the object drawings raised the possibility that a floor effect may have constrained the differences between the groups on that task and, consequently, inflated the critical interaction of groups-by-stimulus type obtained on the full data set. However, the interaction was also obtained on the portion of the data deriving from the second half of the experiment (trials 11 through 20) in which the standard error of the means for the tasks differs very little. Thus, we may suppose that the obtained interaction is genuine and not artificially inflated by a floor effect. It should be noted that these results with second graders parallel those of another recent investigation that demonstrated a material-specific deficit in serial memory in adolescent poor readers (Holmes & McKeever, 1979).

It appears then that poor readers do have a material-specific deficit in memory for order. By way of explanation, two possible alternatives suggest themselves: The deficit may reflect either the ineffective use of phonetic codes or a preference for different and less efficient coding strategies. There is some evidence that poor readers show both types of problems. A recent study (Byrne & Shea, 1979) indicates that, if given a choice, the poor reader does have a preference for an inefficient semantic strategy in retaining linguistic material, but can use a phonetic code, albeit poorly, when no other option is available.

Given the pattern of results obtained in our study, the difficulty of the poor readers could be interpreted as arising from either of the abovementioned causes---the choice of an inappropriate strategy or the inefficient use of the appropriate one. As to the first possibility, the poor readers of the present study may have chosen to use a semantic code, for example, to retain the order of the object drawings, whereas the good readers opted instead for phonetic codes since that is their usual strategy. If this were the case, our data indicate that a semantic coding strategy was certainly inappropriate for the task, since the performance of the poor readers was worse than that of the good readers. The second possibility, which seems to us more likely, is that the requirement of retention of item order may have induced both good and poor readers to attempt to use a phonetic memory strategy, but that the poor readers were less able to do so. Evidence supporting this second possibility is found in several studies in which even poor readers show some susceptibility to phonetic confusion in ordered recall of linguistic material, such as letter strings (Liberman et al., 1977; Shankweiler et al., 1979) or word strings (Mann et al., 1980).

Poor readers, thus, can use a phonetic strategy at times. We must therefore ask what accounts for the greater proficiency of the good readers in tasks, such as ordering the object drawings, where this strategy is clearly both possible and appropriate. An appeal cannot be made to differences in the intelligence of good and poor readers because the pattern of results is

unaltered when the effect of IQ is held constant. It is conceivable that good and poor readers differ in the facility with which they can recode visual stimuli linguistically, and that the poor readers' difficulties may arise in part from slowness in the initial conversion from pictorial to phonetic form. This view receives some support from experiments that indicate that poor readers characteristically take more time than good readers to name a set of recurring items (e.g., color patches) when there is a premium on speed of response (Denckla & Rudel, 1976). However, previous experimental findings of our own (Liberman et al., 1977; Shankweiler et al., 1979) give us reason to believe that the poor readers' problem goes beyond any possible slowness in phonetically recoding a visual stimulus. In those studies, a differential effect for rhyme was found for both good and poor readers in the recall of letters, whether the letters were presented visually as shapes or auditorily as names. Similarly, the Byrne and Shea (1979) study, which involved auditory presentations of stimulus items, also found a deficiency in the poor readers' memory for words and nonwords. Thus, the difficulties of poor readers cannot be due solely to inefficiency in recoding visual stimuli as such. Much the same conclusion was argued on other grounds by Perfetti, Finger, and Hogaboam (1978). We can probably also rule out differences in the rate at which good and poor readers scanned the drawings (Katz & Wicklund, 1971, 1972). In sum, the factors that limit fully effective use of phonetic coding by poor readers have yet to be identified, but some major possibilities can now be eliminated.

With regard to order memory, the present findings are consistent with other indications that children with specific reading disability as a group do not have a general problem in remembering order. Instead, the results suggest that these children do have a general problem in coding information linguistically. In all situations in which phonetic coding would be applicable and desirable, their performance is hampered. In contrast, it is not affected, or less so, when other strategies can be utilized. Insofar as poor readers do have problems with order memory, their difficulties in that domain may be more parsimoniously viewed as further manifestations of their failure to make full use of phonetic coding in working memory.

REFERENCE NOTE

1. Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. *Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability.* Manuscript submitted for publication, 1980.

REFERENCES

Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhardt's framework for memory research. *Psychological Review*, 1978, *85*, 139-152.

Bakker, D. J. *Temporal order in disturbed reading*. Rotterdam, The Netherlands: Rotterdam University, 1972.

Benton, A. L. Developmental dyslexia: Neurological aspects. In W. J. Friedlander (Ed.), *Advances in neurology* (Vol. 7). New York: Raven Press, 1975.

Byrne, B., & Shea, P. Semantic and phonetic memory codes in beginning readers. *Memory & Cognition*, 1979, *7*, 333-338.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N.J.: Erlbaum, 1975.
Comprehensive test of basic skills. Monterey, Calif.: CTB/McGraw-Hill, 1974.
Corkin, S. Serial-ordering deficits in inferior readers. Neuropsychologia, 1974, 12, 347-354.
Crowder, R. G. Language and memory. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic. Cambridge, Mass.: MIT Press, 1978.
Denckla, M. B., & Rudel, R. G. Rapid 'automatized' naming (R. A. N.): Dyslexia differentiated from other learning disabilities. Neuropsychologia, 1976, 14, 471-480.
Holmes, D. R., & McKeever, W. F. Material specific serial memory deficit in adolescent dyslexics. Cortex, 1979, 15, 51-62.
Katz, L., & Wicklund, D. A. Word scanning rate for good and poor readers. Journal of Educational Psychology, 1971, 62, 138-140.
Katz, L., & Wicklund, D. A. Letter scanning rate for good and poor readers in grades two and six. Journal of Educational Psychology, 1972, 63, 363-367.
Kimura, D. Right temporal-lobe damage. Archives of Neurology, 1963, 8, 264-271.
Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conference. Hillsdale, N.J.: Erlbaum, 1977.
Makar, B. W. Primary phonics (Workbooks 1-4). Cambridge, Mass.: Educators Publishing Service, Inc., 1969.
Mann, V. A., Liberman, I. Y., & Shankweiler, D. Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 1980, 8, 329-335.
Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. Phonetic recoding and reading difficulty in beginning readers. Memory & Cognition, 1977, 5, 623-629.
Noelker, R. W., & Schumsky, D. A. Memory for sequence, form, and position as related to the identification of reading retardates. Journal of Educational Psychology, 1973, 64, 22-25.
Perfetti, C. A., Finger, E., & Hogaboam, T. Sources of vocalization latency differences between skilled and less skilled young readers. Journal of Educational Psychology, 1978, 70, 730-739.
Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 531-545.
Slosson, R. L. Slosson intelligence test for children and adults. East Aurora, N.Y.: Slosson Educational Publications, 1963.
Stanley, G., Kaplan, I., & Poole, C. Cognitive and nonverbal perceptual processing in dyslexics. Journal of General Psychology, 1975, 93, 67-72.
Vellutino, F. R., Pruzek, R. M., Steger, J. A., & Meshoulam, U. Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. Cortex, 1973, 9, 368-384.
Vellutino, F. R., Steger, J. A., & Kandel, G. Reading disability: An investigation of the perceptual deficit hypothesis. Cortex, 1972, 8, 106-118.
Woodcock, R. W. Woodcock reading mastery tests. Circle Pines, Minn.: American Guidance Service, Inc., 1973.

PERCEPTUAL EQUIVALENCE OF TWO KINDS OF AMBIGUOUS SPEECH STIMULI*

Bruno H. Repp

Abstract. Stimuli from two synthetic /da/-/ga/ continua were presented in a speeded labeling task. One continuum was generated by parameter interpolation; the other, by adding the waveforms of the endpoint stimuli in varying proportions. Both continua showed an increase in latencies at the category boundary, suggesting that the two procedures yield equally ambiguous stimuli.

Ambiguous stimuli play a central role in speech perception research. By virtue of their perceptual instability, they serve as indicators of a large variety of laboratory phenomena, including categorical perception, selective adaptation, phonetic trading relations, and all sorts of context effects. Traditionally, ambiguous stimuli have been constructed with the aid of speech synthesizers: Two unambiguous stimuli from different phonetic categories are selected, and a number of steps are interpolated between their parameter values, leading to a continuum that includes some ambiguous stimuli in the region of the phonetic category boundary. Until recently, this was the only method available. However, a new technique was applied in a recent doctoral thesis by Stevenson (1979). Instead of interpolating parameter values between two endpoint stimuli, he added the digitized waveforms of the endpoint stimuli in various proportions, increasing the amplitude of one component waveform while decreasing that of the other and so producing a continuum. In fact, he was able to construct such continua from carefully aligned natural utterances of /ba/, /da/, and /ga/; but the technique can, of course, be used with synthetic speech as well.

Electronically mixed synthetic stimuli have been used previously, primarily to compare their perception with that of the same component stimuli presented dichotically (Halwes, 1969; Porter & Whittaker, 1980; Repp, 1976, 1980). However, Stevenson (1979) was apparently the first to construct whole stimulus continua that way. His technique is interesting, especially because it can be used with natural speech. However, are there any important perceptual differences between an ambiguous stimulus created by superimposing two unambiguous stimuli and one characterized by a single set of intermediate parameters? Stevenson used his stimuli in a variety of standard experimental tasks, including categorical perception, selective adaptation, and dichotic listening, and obtained results very similar to those found with traditional stimulus continua, although he never performed any direct comparison.[1]

---

*To be published in the Bulletin of the Psychonomic Society.
Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I am grateful to Dale Stevenson for bringing his work to my attention and for commenting on an earlier version of this paper.

The present study explored one way in which the two types of ambiguous speech stimuli might differ in perception. When presented with an ambiguous stimulus of the traditional kind, which has acoustic properties that are truly intermediate, listeners experience uncertainty that increases the time needed to assign the stimulus to one of two categories (Studdert-Kennedy, Liberman, & Stevens, 1963; Pisoni & Tash, 1974). However, when listening to a stimulus from a Stevenson continuum, which contains two unambiguous sets of cues superimposed, there might be no uncertainty on a given trial; rather, perception might go with one or the other set of unambiguous cues on a probabilistic basis. The present study tested this hypothesis by examining whether the characteristic peak in identification latencies at the category boundary of traditional speech continua (Studdert-Kennedy et al., 1963; Pisoni & Tash, 1974) is present to the same extent on a continuum of electronically mixed stimuli.

Method

Subjects. Eight paid student volunteers participated. They had little or no experience in experiments of this kind.

Stimuli. The syllables /da/ and /ga/ were synthesized on the OVE IIIc synthesizer at Haskins Laboratories. They were distinguished only by the third-formant (F3) transition whose onset frequency was 2976 Hz in /da/ and 2150 Hz in /ga/. All other characteristics were shared: fully periodic waveform, a duration of 250 msec, a fundamental frequency that fell linearly from 110 to 80 Hz, 50-msec linear formant transitions, F1 rising from 285 to 771 Hz, F2 falling from 1770 to 1233 Hz, and an F3 steady-state frequency of 2520 Hz.

The mixed (Stevenson-style) continuum was constructed in the following way: The two syllables were digitized at 10 kHz using the Haskins Laboratories PCM system. Nine intermediate stimuli were obtained by adding the /da/ and /ga/ waveforms point by point after reducing the amplitude of each by a certain amount. That amount was determined by translating the ratios 1:9, 2:8, ... 8:2, 9:1 into dB values under the constraint that the amplitude of the combined waveforms remain constant. The resulting attenuation values were -1, -2, -3, -5, -6, -8, -10, -14, and -20 dB SPL for the /da/ component; they applied in inverse order to the /ga/ component.[2] Only these nine stimuli were used in the experiment.

The interpolated (traditional) continuum was constructed by synthesizing eight intermediate stimuli between /da/ and /ga/, changing the onset frequency of F3 in equal decrements. All ten stimuli were digitized at 10 kHz. To control for any possible artifacts due to waveform addition on the other continuum, and to match the numbers of stimuli on the two continua, the ten stimuli were reduced to nine by adding the waveforms of neighbors on the continuum. Stimulus amplitudes were first reduced by 6 dB SPL, to match the amplitudes of the stimuli on the mixed continuum.

Randomized stimulus sequences were recorded on tape. The stimuli from both continua were randomized together to yield a basic unit of 18 stimuli. Five such units formed one continuous block of 90 stimuli, with interstimulus intervals of 2 sec. Four such blocks were recorded, with longer pauses in

between. Each block was prefixed with four warm-up stimuli that were not scored. At the very beginning of the tape was a practice sequence of 40 stimuli containing only instances of the endpoint stimuli of the two continua. To the author, the stimuli from the two continua were phenomenally indistinguishable.

Procedure. Subjects were tested individually in a soundproof booth. They sat in front of a table and rested their index fingers on two telegraph keys labeled "dah" and "gah". The response-to-keys assignment was counterbalanced across subjects. The instructions stressed speed of response. The subjects were permitted to stop the tape recorder by remote control between blocks and take a rest, if desired. The tape was played back on a Crown 800 tape recorder located in an adjacent room, and the subject listened over Telephonics TDH-39 earphones. Reaction times were measured by a Hewlett-Packard 5302A 50MHz universal counter and printed out by a Hewlett-Packard 5150A thermal printer. The counter was triggered by a signal recorded on the second tape channel and synchronized with syllable onset.

Results and Discussion

The results, averaged over subjects, are displayed in Figure 1. It can be seen that the labeling functions for the two continua were virtually identical, and so were the latency functions. The, perhaps fortuitous, coincidence of the category boundaries[3] is less important than the fact that both latency functions exhibited peaks of equal magnitude at the category boundary. Analysis of variance confirmed a significant effect of stimulus number, $F(8,56) = 2.85$, $p < .01$, but no significant effect involving type of continuum.

Thus, the two kinds of continua were perceptually equivalent in this speeded labeling task. In particular, stimuli from the centers of the two continua were equally ambiguous and created equal uncertainty in listeners. This tells us something about the perceptual processing of mixed stimuli. Apparently, it is not the case that the superimposed conflicting cues are accessed individually by some selective attention mechanism (as perhaps suggested by the concept of auditory "listening bands"--Divenyi, 1979) or subject to mutual lateral inhibition or masking. Rather, conflicting transitions of the same formant seem to engage in a "trading relation," just as transitions of different formants do (see Mattingly & Levitt, 1980, for a recent study). The outcome of this trade-off appears to be perceptually equivalent to an acoustically intermediate specification, at least as far as phonetic perception is concerned. Stevenson's (1979) extensive data obtained with electronically mixed stimuli suggest that they are equivalent to traditional stimuli in many other respects. It seems unlikely, then, that the new technique of stimulus construction will lead to any new insights about the mechanisms of speech perception, although it deserves continued attention because of its applicability to natural speech.

Several limitations of Stevenson's method should be pointed out, however. First, it can be used only with stimuli of similar temporal structure, i.e., it is restricted primarily to variations in spectral cues (see also Footnote 1). Second, it does not work with stimuli that do not readily fuse into a single percept, such as vowels (Stevenson, 1979). The factors at work here

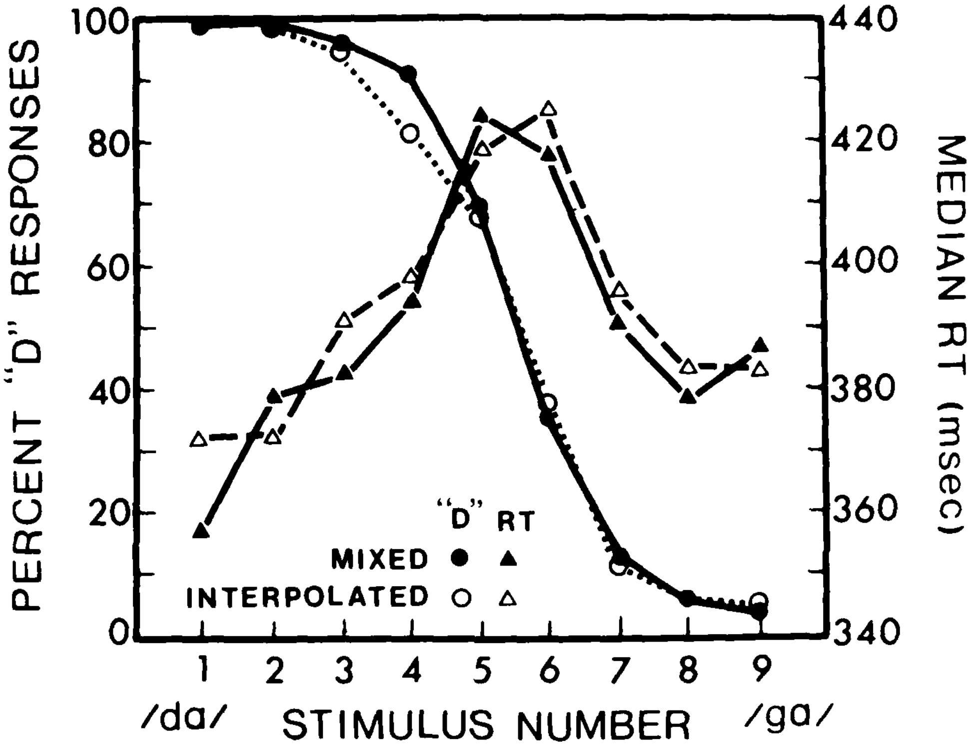


Figure 1. Identification ("d" responses) and latency functions for two kinds of /da/-/ga/ continua.

seem to be very similar to those governing dichotic fusion (cf. Cutting, 1976). Third, mixed continua have the property that stimuli become increasingly less discriminable (on purely auditory grounds) the farther they are from the center of the continuum, which is undesirable in categorical-perception experiments, where the detectability of within-category differences is of prime interest. Therefore, it appears that Stevenson's technique will be useful only under very special circumstances.[4]

REFERENCES

Cutting, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. Psychological Review, 1976, 83, 114-140.

Divenyi, P. L. Some psychoacoustic factors in phonetic analysis. Proceedings of the Ninth International Congress of Phonetic Sciences, Vol. II. Copenhagen: University of Copenhagen, 1979, 445-452.

Halwes, T. G. Effects of dichotic fusion on the perception of speech. Unpublished doctoral dissertation, University of Minnesota, 1969.

Mattingly, I. G., & Levitt, A. G. Perception of stop consonants before low unrounded vowels. Haskins Laboratories Status Report on Speech Research, 1980, SR-61, 167-173.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics, 1974, 15, 285-290.

Porter, R. J., Jr., & Whittaker, R. G. Dichotic and monotic masking of CV's by CV second formants with different transition starting values. Journal of the Acoustical Society of America, 1980, 67, 1772-1780.

Repp, B. H. Identification of dichotic fusions. Journal of the Acoustical Society of America, 1976, 60, 456-469.

Repp, B. H. Stimulus dominance in fused dichotic syllables: Trouble for the category goodness hypothesis. Journal of the Acoustical Society of America, 1980, 67, 288-305.

Stevenson, D. C. Categorical perception and selective adaptation phenomena in speech. Unpublished doctoral dissertation, University of Alberta, 1979.

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. Journal of the Acoustical Society of America, 1963, 35, 1900. (Abstract)

FOOTNOTES

[1]Stevenson (1979) drew an analogy between his ambiguous stimuli and certain ambiguous visual figures, such as the Necker cube: A continuum can be constructed by beginning with an unambiguous drawing of orientation A of the (opaque) cube and by then slowly increasing the intensity of the added line segments unique to orientation B while decreasing the intensity of the line segments unique to A until only B remains. At the center of the continuum, where all lines are equally intense, we have the maximally ambiguous figure--the (transparent) Necker cube. It is interesting to note that this visual analogy is not appropriate for the traditional method of constructing speech continua; if applied to the cube drawings, that method would use spatial interpolation between lines unique to the two endpoint stimuli, resulting in curvilinear distortions that destroy the identity and three-dimensionality of the cube. However, the interpolation technique could be used to construct a

continuum from, say, a circle to a square, whereas Stevenson's method would fail here because intermediate stages would be seen as a square superimposed on a circle, not as one or the other. Apparently, the endpoint stimuli must have a rather special relation to each other if both methods shall result in truly ambiguous stimuli. It appears that this condition is satisfied only by certain speech stimuli, such as stop-consonant-vowel syllables differing in (stop) place of articulation.

[2]Since only integer dB values could be used on the computer, overall amplitude varied over a range of 0.5 dB SPL. Also, the calculated values strictly apply only to perfectly correlated waveforms (cf. Stevenson, 1979). However, since the present stimuli differed only in F3, and only during the first 50 msec, the values used were quite adequate.

[3]The author, as a pilot subject, had different boundaries on the two continua. No claim is being made here that the two continua constitute equivalent perceptual scales, i.e., that there is a one-to-one equivalence of stimuli.

[4]This conclusion is not intended as a critique of Stevenson whose careful and sophisticated (but, unfortunately, unpublished) work made a valuable methodological contribution.

PRODUCING RELATIVELY UNFAMILIAR SPEECH GESTURES:
A SYNTHESIS OF PERCEPTUAL TARGETS AND PRODUCTION RULES

G. J. Borden,* K. S. Harris,** Hollis Fitch,*** and H. Yoshioka****

Abstract. Attempts of speakers to imitate familiar and foreign syllables under adverse feedback conditions were analyzed by perceptual judgments, electromyographic recordings, and spectrographic measures. Although foreign syllables were more poorly imitated than familiar syllables, decrements in feedback interfered more with familiar than with novel utterances. Decrements in acoustic, tactile, and proprioceptive information were worse in combination than singly. Speakers did not improve unfamiliar fricative production under any condition on 13 learning trials.

Research during the last decade has demonstrated that intelligibility of the speech of skilled speakers remains high despite masking of the speakers' auditory feedback or decreasing their tactile feedback. There is some segmental distortion when the tongue is anesthetized (Ringel & Steer, 1963; Borden, Harris, & Oliver, 1973), and some prosodic distortion when speech is attempted under simultaneous but modified auditory feedback (Lane & Tranel, 1971; Siegel & Pick, 1974), but the overall effect upon speech production seems to be surprisingly small (see Borden, 1979, for a review).

These findings argue for the importance of a feedforward system for production of well-known motor patterns for speech, with auditory and tactile information used for fine tuning or correction of errors. The adult speaker seems to know the possibilities of his or her own vocal tract. Simple constraints on movement imposed by talking with a pipe or pencil clenched between the teeth or with an experimental bite block do not alter the vocal tract dimensions, and interference with speech is minimal (Lindblom & Sundberg, 1971). These results are consonant with those of animal experiments, in which direct and complete elimination of sensory information is accomplished by surgical means. It has been shown that monkeys trained to perform specific movements can continue to do so, despite deprivation of feedback from limbs or chewing muscles (Taub & Berman, 1968; Goodwin & Luschei, 1974; Polit & Bizzi, 1978), although there are indications that new movements are impaired (Polit & Bizzi, 1978).

---

*Also Temple University, Philadelphia, PA.
**Also City University of New York, New York, NY.
***Also Institute for Defense Analyses, Princeton, NJ.
****Also University of Tokyo, Tokyo, Japan.
Acknowledgment. The writing of this article was partially supported by NIH grant NS13870 to Haskins Laboratories. We are grateful to Arthur Abramson for taping the model stimuli and for judging the subjects' utterances under normal conditions, and to Anders Löfqvist for serving as the expert listener.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

Of course, we cannot know with certainty the role of self-monitoring of speech, because it is impossible to eliminate simultaneously all channels of information available to a speaker. We do know that sensory information is important in learning speech for the first time or in successfully learning the speech patterns of a new language. The labored speech of the deaf (Osberger & Levitt, 1979; Harris & McGarr, 1980) and the rare case of a speaker with an oro-sensory loss (Chase, 1967) testify to the importance of self-monitoring while learning speech. We know, too, that normal adult speakers often need time to adjust to prosthetic devices that alter the dimensions of their vocal tracts (Hamlet & Stone, 1976), and feedback of auditory, tactile, and proprioceptive information is presumed to control the compensatory patterns that evolve.

There have been no studies to our knowledge, however, that investigate self-monitoring of speech by comparing the effects of diminished sensory information on the performance of speech gestures new to the speaker with those familiar to the speaker, with the exception of one report to the effect that children who are better than other children at identifying forms placed in the mouth (oral stereognosis) are also better at learning non-native speech sounds (Locke, 1968).

In the present investigation, we were interested in exploring whether perceptually accurate speech sounds would be produced under conditions of adverse speech control when the speech gestures were not those learned as part of the language of the speaker. How well might the speaker control production of non-English syllables? Might vowels and consonants depend differently upon sensory information? How well might the speaker control English and non-English utterances when auditory feedback is diminished? when tactile information is decreased? when vocal tract configuration is altered? The question that motivated these experiments was not what happens to speakers with loss of feedback—the speaking conditions reported in this paper represent diminished or altered feedback, not its absence—rather, the question is how do familiar versus relatively novel speech gestures hold up under various conditions and combinations of conditions that alter or diminish information that is normally fed back to the speaker as he is talking?

Two approaches can be taken to judge adequacy of performance. One approach is to measure some aspect of production directly in various conditions—here we have measured articulator activity using EMG techniques, and some aspects of acoustic output, using conventional spectrographic analysis. Another approach is to examine perceptual adequacy by using listener judgments of performance. The second approach has the disadvantage of being subjective, but does measure communicative adequacy. While the first approach is objective, any particular set of measurements is not exhaustive.

One can rationalize three hypotheses about the experimental outcome: The first is that relatively novel utterances will suffer more than familiar utterances under conditions of altered or diminished information, because speakers might need more information for the less familiar utterances. The second hypothesis is that familiar utterances would suffer more than novel under deprived feedback conditions, because speakers may hold internalized finely developed auditory-oro-sensory criteria for the well-learned utterances and might use feedback to sharpen the match between their utterances and these

criteria. For the less familiar utterances, however, speakers may hold only broad criteria for how the speech should sound and feel, and therefore make less use of information from the periphery. The third hypothesis is that familiarity would make little difference, because speakers might not succeed in producing unfamiliar motor sequences even when all feedback information is available; they might convert less familiar utterances into familiar ones and utter a variant of a similar sound from their own language system.

PROCEDURE: PRODUCTION TASK AND PERCEPTUAL ANALYSIS

The general design of the investigation was to have subjects imitate a recording of a phonetician saying syllables that were within the phonetic inventory of English and syllables that were phonetically foreign to English. The speakers imitated the speech sounds under normal speaking conditions and under altered speaking conditions: auditory masking, lingual anesthesia, and alterations of the shape of the palatal vault. The speech was recorded acoustically and the muscle activity of the tongue was analyzed by electromyographic measures. Tapes for each speaker made by pairing utterances spoken by the phonetician with utterances spoken by a subject under various speaking conditions were used for perceptual tests to assess the judged differences between speaking conditions.

Subjects

Three normal adult males served as the primary subjects for the experiment. They were speakers of American English, and, although they had studied languages other than English in school, each subject was essentially monolingual with little practical experience in speaking any other languages. Two of the subjects were 21 years old (DB and GF) and the third was 33 (TB). None was informed of the purpose of the experiment. Since the long-lasting effects of the anesthesia condition precluded the perfect balancing of orders, four additional subjects were recorded each with a different order of conditions. These speakers were run without nerve-block anesthesia of the tongue and without electromyographic insertions to see what order effects there might be, and to enlarge the subject pool. The non-nerve-block speakers were students at Temple University and were also naive about the purpose of the investigation. As other subjects were used for the perceptual part of the analysis, we shall avoid confusion by referring to the imitators as speakers and to the subjects of the perceptual tests as listeners.

Speech Task

For this investigation we chose a small set of speech sounds, some that would be familiar to monolingual speakers of American English and some that would be relatively novel. The criteria were that the sounds must exist in some language and they must be acoustically distinct. We chose two familiar vowels [i] and [eᶦ] (as in 'see' and 'say') and two familiar consonants, one voiceless [ʃ] and one voiced [z] (as in 'shoe' and 'zoo'). For the less familiar sounds the vowels [y] and [ø] (as in the French words 'tu' and 'deux') were chosen because they are rounded front vowels not phonologically

present in English. The novel consonants chosen were the voiceless and voiced palato-velar fricatives [x] and [ɣ] (as in the Spanish words 'rojo' and 'rogar'). The vowels were initiated with [p] and the fricatives were followed by [i] yielding eight syllables. A phonetician proficient in the production of all eight of these speech sounds recorded them in syllable form after the word 'say.' The list was read three times and a satisfactory token of each type was chosen to be digitized on the Haskins PCM system. A tape recording was constructed containing a list of 24 utterances (each utterance repeated three times and randomized) followed by eight lists in which each syllable type was repeated 10 times. The last eight repetition lists were used to investigate learning.

Experimental Conditions

The three primary speakers from whom electromyographic data were collected were recorded under conditions of auditory masking, lingual anesthesia, false palate, and combinations of these conditions as well as the normal speaking condition used as a control. The four speakers recorded without EMG insertions were recorded in the same conditions as the primary subjects with the exception of the condition of lingual anesthesia.

The condition of diminished auditory feedback was achieved by recording the speech of the phonetician on one channel and white noise on the second channel of a tape recording. The speech was delivered binaurally at 70 dB SPL and the white noise, also binaurally, at 90 dB SPL during the subjects' responses. To control vocal intensity, subjects were instructed to monitor the VU meter on the tape recorder that was recording their responses: they were not to let their vocal intensity rise above the midpoint of the range, representing about 55 or 60 dB. Although the low frequencies of the voice were undoubtedly transmitted by bone conduction, the higher frequency contribution of the vocal tract resonances to the various speech sounds was essentially masked for the speaker.

Lingual anesthesia was produced by blocking the sensory fibers of the lingual nerve on both sides of the jaw. The lingual nerve, a branch of the Trigeminal nerve, was blocked by a dentist who bilaterally injected 1.8 cc of 3 percent Carbocaine containing a vasoconstrictor. The criterion for lingual anesthesia was that the superior surface of the anterior two-thirds of the tongue must be insensitive to a dental probe.

The conditions of masking noise and nerve block resulted in diminished auditory and tactile feedback, respectively. Proprioceptive feedback, in this case information on tongue position and movement, is impossible to interrupt short of surgical techniques. To impoverish the usefulness of the proprioceptive information, however, the shape of the vocal tract was altered by placing a dental impression material, Alginate, on the superior alveolar ridge behind the central and lateral incisors. The material extended posteriorly along the hard palate for several centimeters. Whatever proprioceptive information the speaker may have received from tongue position and movement within the vocal tract, the fact that vocal tract volume was changed, thus altering the presumed coordinates of the space, would alter the customary reference points for proprioceptive information. After the impression material was removed

from the mouth of each subject, it was cut along a central line extending from between the central incisors, and the width of the portion corresponding to the apex of the alveolar ridge was measured. The addition of material resulted in a buildup of the ridge by 6 mm for each subject.

Conditions were applied singly and in combination with orders varied. For the primary speakers, speaking conditions were given in the following order. M stands for auditory masking, NB for the nerve block resulting in lingual anesthesia, and A for Alginate, the dental impression material used to alter the architecture of the palate. For DB, the order was NB, M, and A; NB and A; and finally, NB alone. For TB, the order was NB; NB and M; NB, M, and A, and NB and A. For GF, the order was M; M and A; A; and M, A, and NB. The control condition was recorded on another day to ensure that there were no effects of the anesthesia. For the non-nerve-block subjects four orders were possible reserving the control condition for last: 1) A; A and M; M, 2) M; A; A and M, 3) M; A and M; A, and 4) A and M; A; M. The order A and M; M; A was not possible as the impression material could not be removed and reinserted.

Electromyographic Recording

Hooked wire electrodes of .002 inch platinum alloy were inserted into the superior orbicularis oris muscle (OO), the superior longitudinal muscle (SL), the inferior longitudinal muscle (IL), and the genioglossus muscle (GG) of the three primary subjects. The orbicularis oris muscle was sampled to allow for observations of muscle activity in the lips for the rounded less familiar vowels. The genioglossus muscle was sampled to assess production of the high front vowels, and the intrinsic muscles of the tongue were sampled in an effort to observe differential tongue activity for production of the the fricatives. The EMG recordings consisted of the eight speech task utterance types, 13 tokens of each, under all speaking conditions. Only the EMG signals recorded during the three tokens of each type in the randomized list of 24 items have been analyzed. The signals were rectified, smoothed with a 35 msec time constant, and digitized. Procedures for insertion, recording, and analysis are described in detail elsewhere (Hirose, 1971; Kewley-Port, 1973).

Acoustic Recording

Sound spectrograms were made of all utterances spoken by the primary speakers. Second formant frequencies were measured for [i], [eɪ], [y] and [ɸ]. Normally, the rounded front vowels /y/ and /ø/ are realized acoustically with higher F_1 and lower F_2 than the unrounded front vowels /i/ and /eɪ/ (Pols, Tromp, & Plomp, 1973). The tongue is thought to be higher for the unrounded members of the respective pairs /i-y/ and /eɪ-ø/ (Raphael, Bell-Berti, Collier, & Baer, 1979).

The fricative consonants were measured in the center of the third formant noise. Normally, the prominent resonances for [ʃ] are lower in frequency (approximately 2500 Hz) than those for [z] (approximately 4000 Hz). Figure 1 contrasts the [ʃ] and [z] resonances in the model "Say [i]" and "Say [zi]." Figure 2 shows the spectrographic representation of the utterance "Say [Xi]" and "Say [ɣi]" as spoken by the phonetician used in this study. F_2 and

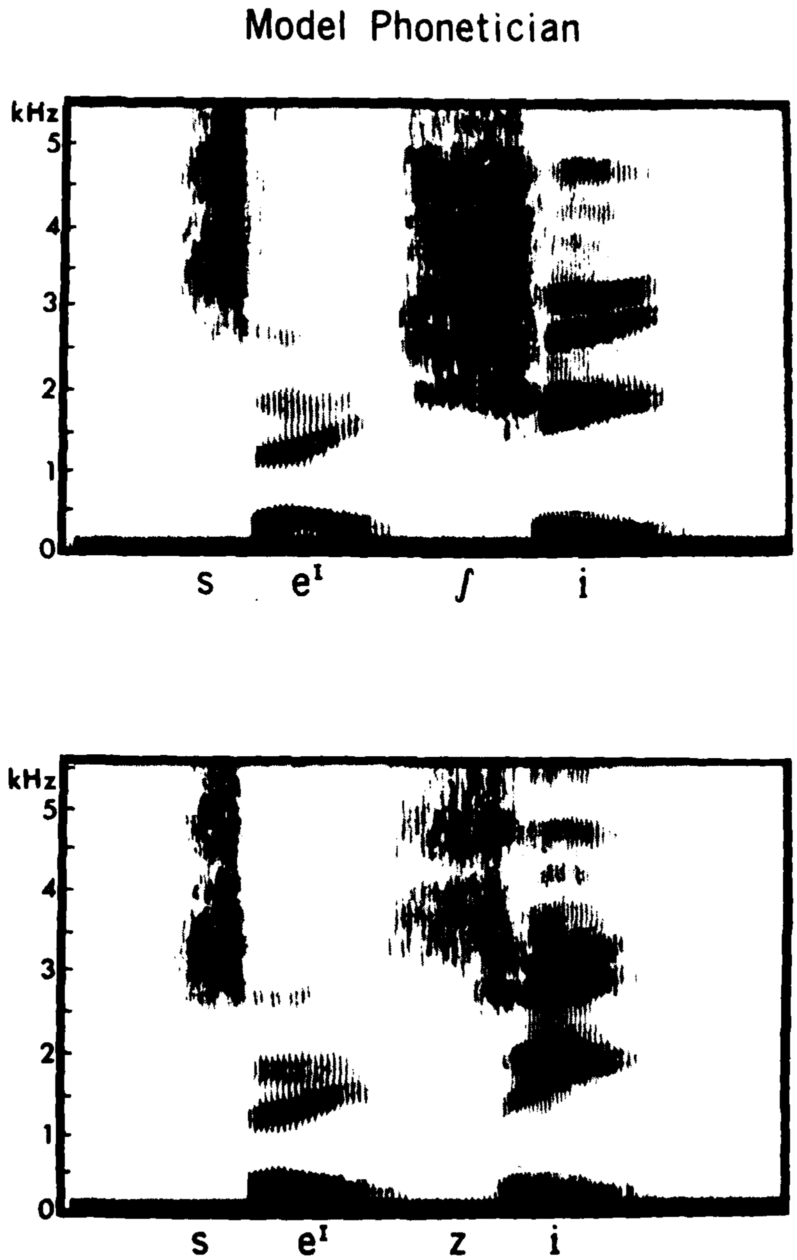


Figure 1

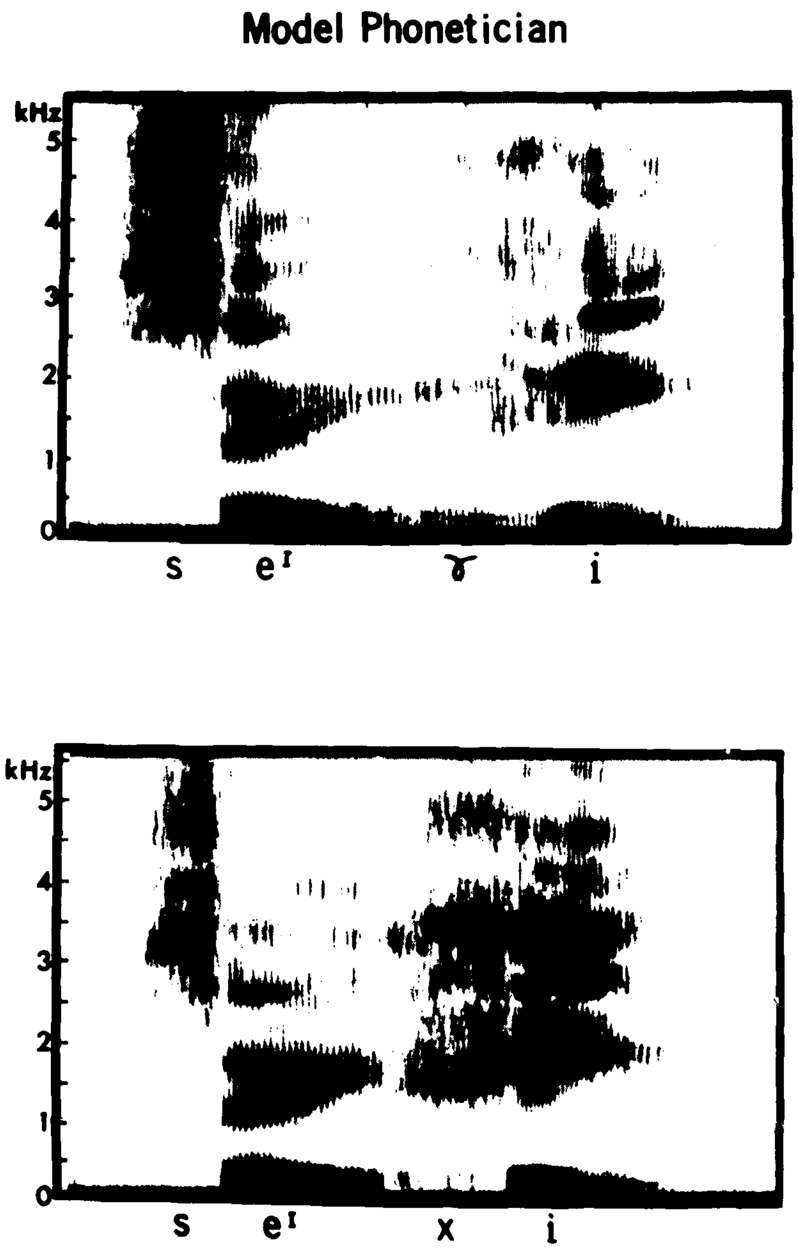
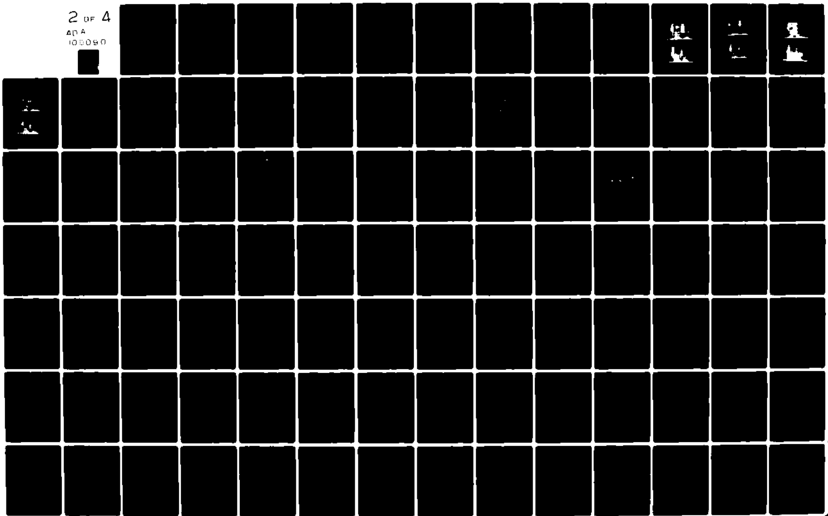


Figure 2

AD-A105 090 HASKINS LABS INC NEW HAVEN CT F/G 17/2
STATUS REPORT ON SPEECH RESEARCH, 1 APRIL-30 JUNE 1981.(U)
1981 PHS-N01-HD-1-2420
UNCLASSIFIED SR-66(1981) NL

AD A
105090



F_3 are close together for [X] and [ɣ] with F_3 ranging between 2000 Hz and 2500 Hz in an average male vocal tract. Conspicuous is the antiresonance below the second formant. In cases where fricative energy was low in the F_3 region, the F_3 frequency at the onset of the following vowel was measured.

Perceptual Testing

A listening tape was constructed from the model utterances of the phonetician and the imitations of each subject (one tape for each speaker) by digitizing all of the speech samples and editing them on the Haskins PCM system so that for each syllable type, each speaking condition was contrasted with each of the other speaking conditions in both orders. Each trial presented the model utterance, for example "Say zi" as said by the phonetician, followed by the speaker's imitation under one condition, then the phonetician again, followed by the speaker's imitation under another condition. The phonetician's utterance and the imitations were separated by 500 msec, and the pairs for each trial were separated by 1500 msec. A 3 second pause between trials allowed time for listeners to check on answer sheets the imitation they preferred. With five conditions (yielding 10 condition contrasts and with orders of pairs reversed, 20 condition contrasts) of 24 utterances (8 types, 3 tokens each), each listening test consisted of 24 lists of 20 trials each for a test of 480 items. Trials were randomized throughout each test, and each condition was paired with every other condition with orders reversed. Each test was divided into two tapes. Listeners were 27 students from the University of Connecticut, 9 to judge the two test tapes for one of the three speakers. Each tape took approximately one hour. Listeners were asked to judge pronunciation and to disregard any change in loudness or pitch. They were to indicate which of the two imitations in each contrast "more successfully matched the speech sounds" of the phonetician. For the tape constructed from the responses of the first speaker, judgments of three expert listeners were collected to compare with the judgments of the relatively naive student listeners, to assess the effects of listener perceptual sophistication.

Listening tapes were also constructed from the responses of four speakers who did not receive a nerve block. Again, students from the University of Connecticut served as listeners. The listeners were instructed to mark the imitation judged worse than the other with a check and, if much worse, with an X. This change in procedure was an attempt to obtain an idea of the relative magnitude of decrement in perceived pronunciation resulting from the experimental conditions.

RESULTS

Analysis of the data can be divided into the electromyographic analysis, spectrographic analysis, and the analysis of listener judgments. We shall briefly mention the EMG and spectrographic results first, and devote more space to the perceptual results.

Electromyographic Analysis

The first three samples of each syllable type, spoken under each condition, have been analyzed for the three primary speakers. Peak amplitude measures for each electrode placement were graphed. Timing measures were also made.

The muscle that is the primary contributor to lip closure and lip rounding, the orbicularis oris muscle (OO), was active for all three speakers during the rounded vowels [y] and [ø], while it was inactive for [i] and [eᴵ]. Figure 3 shows the contrast between the two types. There is a compact peak of activity for the [p] in [pi] starting before the vertical line at zero. The line indicates the termination of the vowel in 'say' for the utterance 'Say [pi].' The [p] for [py] is also preceded by a compact burst of muscle activity, but OO remains fairly active (324 μv at around 400 msec) throughout the vowel. All these speakers showed evidence of OO activity for the unfamiliar vowels [py] and [pø].

Successful recordings were made from GG for two speakers, and were examined for productions of [i]. Activity was remarkably stable for TB, especially the timing of the activity (Figure 4). Peak amplitude was lower with the addition of Alginate. For speaker GF, GG activity for [i] tended to be more diffuse and drawn out as the speaking conditions got more complicated.

The patterns of activity for SL and IL differed from subject to subject. In general, when either muscle was active for a given fricative, the activity often became erratic with the application of Alginate to the palate, with an increase in activity recorded from IL for two of the speakers. IL normally depresses the tip of the tongue. One speaker (TB) showed little change in SL for [z] in the Alginate condition, but showed a decrease of IL activity (Figure 5). Since TB produces [s] and [z] with the tip of the tongue curled down behind the lower incisors bunching the dorsum of the tongue for the constriction (Borden & Gay, 1979), we assume that the pattern represents a decrease in bunching.

Only one speaker (DB) used SL for fricatives other than [z], limiting comparisons between novel and familiar consonants to that speaker. Comparing the electromyograms of [z], the least variable fricative for DB, with the most variable, [X], activity recorded from SL in the worst speaking condition (nerve-block, alginate, and auditory masking) remained essentially the same for the tokens of [z] but varied considerably for tokens of [X] (Figure 6). The first utterance was transcribed as [z] in all instances, but the second utterance was transcribed as [Xʳ], a velar fricative with a retroflexed tongue, in the first imitation and as [ʕ], a voiced pharyngeal fricative, for the second imitation. SL was active for the better imitation but completely inactive for the pharyngeal fricative.

In general, electromyographic recordings confirmed the observations described below. First, speakers imitated the unfamiliar rounded front vowels adequately in all conditions, using lip rounding to do so. Second, adverse speaking conditions tended to result in reduced tongue activity or erratic patterns.

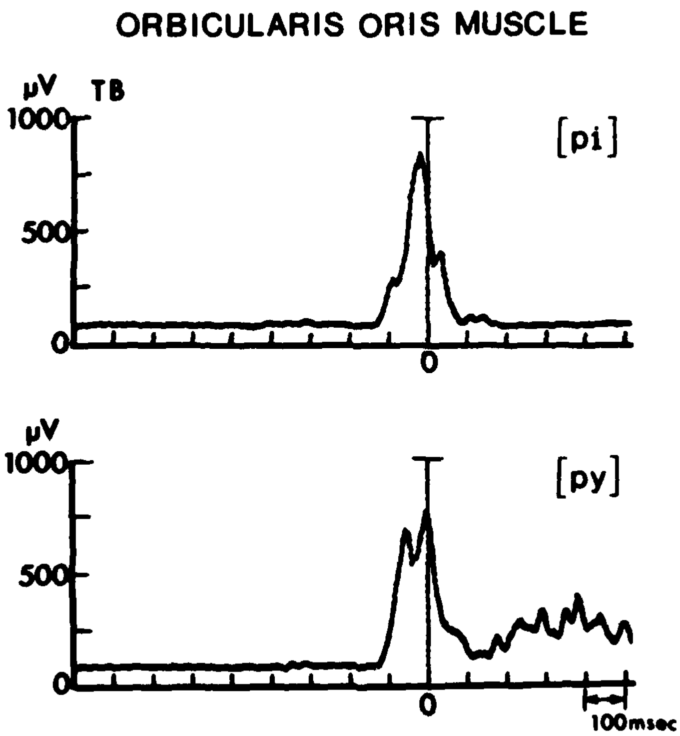

ORBICULARIS ORIS MUSCLE
µV
TB
1000
500
0
0
[pi]
µV
1000
500
0
0
[py]
100msec

Figure 3

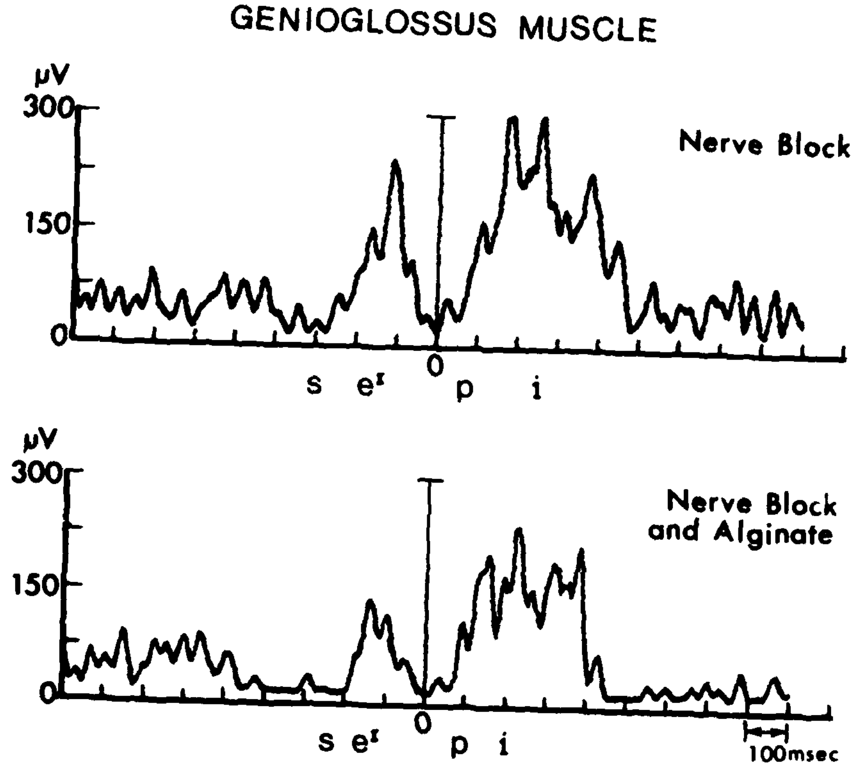

GENIOGLOSSUS MUSCLE
μV
300
150
0
Nerve Block
s eɪ 0 p i
μV
300
150
0
Nerve Block
and Alginate
s eɪ 0 p i
100msec

Figure 4

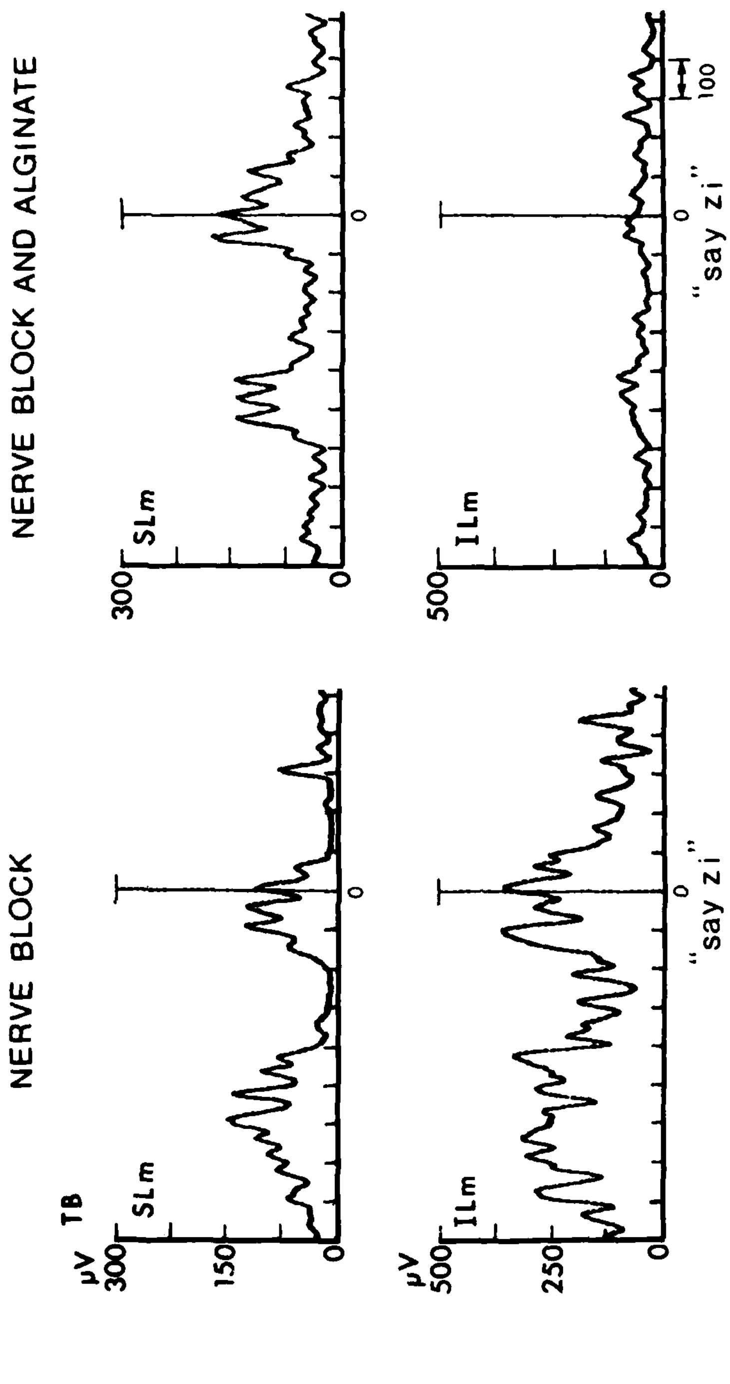

NERVE BLOCK
NERVE BLOCK AND ALGINATE
μV
TB
SLm
ILm
300
150
0
500
250
0
"say zi"
100

Figure 5

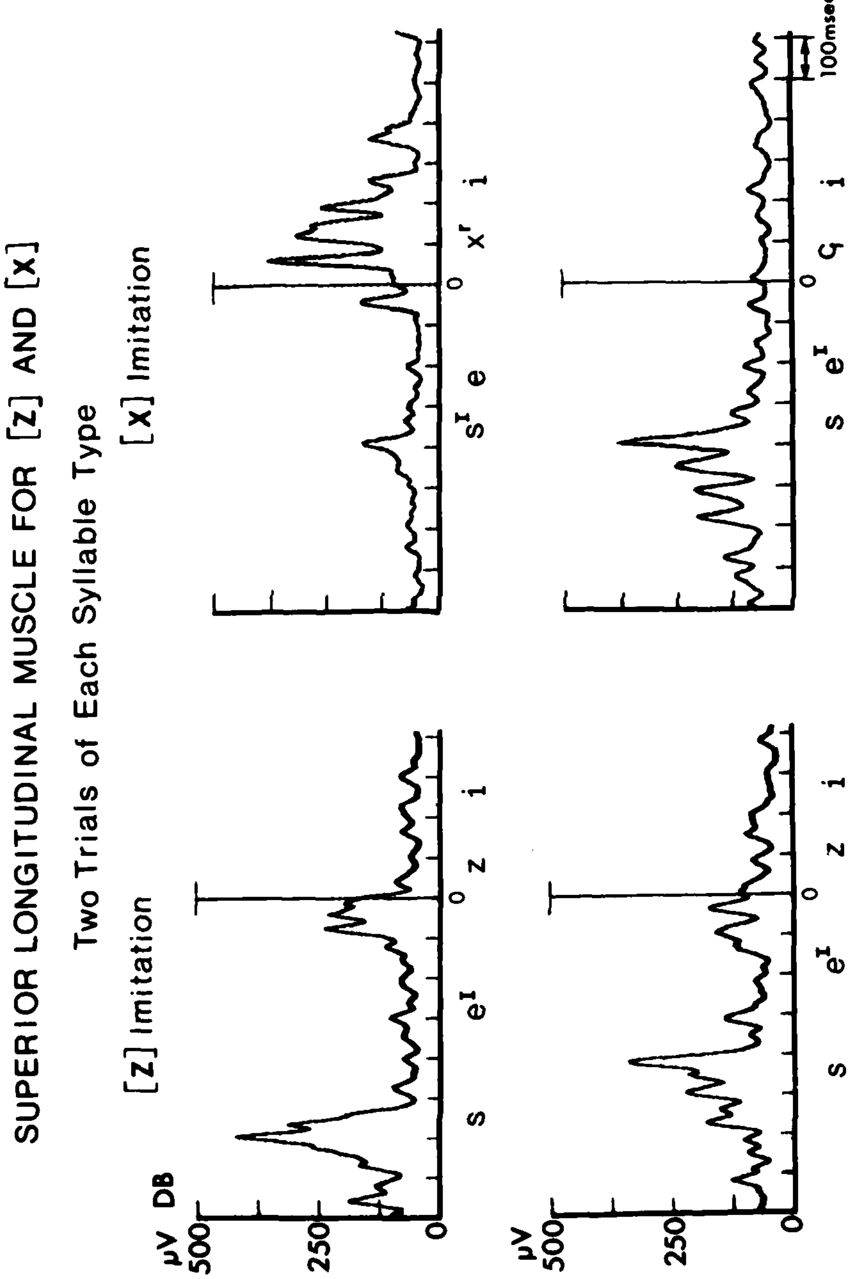

SUPERIOR LONGITUDINAL MUSCLE FOR [Z] AND [X]
Two Trials of Each Syllable Type
[Z] Imitation
[X] Imitation
μV
DB
500
250
0
100msec

Figure 6

Table 1

F_2 for Vowels

S:DB	i		y		e^{I}		ϕ	
	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
N	2233	47	1683	62	1800	0	1533	94
NB	2133	24	1817	131	1867	47	1617	103
NB + M	1900	41	1600	141	1708	12	1517	184
NB, M, & A	1867	47	1733	262	1650	41	1433	94

S:TB	i		y		e^{I}		ϕ	
N	1783	24	1300	41	1400	0	1300	0
NB	1767	23	1416	62	1433	47	1450	122
NB + M	1700	0	1400	62	1283	24	1367	47
NB + A	1700	41	1325	25	1325	25	1266	47
NB, M, & A	1600	71	1233	47	1200	0	1216	23

S:GF	i		y		e^{I}		ϕ	
N	1967	47	1683	24	1717	23	1475	54
M	1733	24	1467	112	1300	20	1383	83
A	1858	24	1683	43	1675	0	1458	83
M & A	1667	23	1383	24	1358	24	1292	59
NB, M, & A	1717	47	1558	83	1500	0	1300	41

Spectographic Analysis

Measurements of the second formants of the vowels for the three primary speakers in this study are presented in Table 1. The means are based on the three tokens of each syllable under each speaking condition that was used in the perceptual test. The acoustic difference between pairs found in previous studies (Pols et al., 1973; Raphael et al., 1979) holds under all speaking conditions for the [i]-[y] contrast: F_2 for [i] is higher than for [y]. The difference is maintained for [eᴵ] and [ø] under normal speaking conditions, but does not hold under all adverse conditions. A prominent effect upon the formant frequencies of the vowels is seen in the condition of auditory masking. Generally, when subjects are prevented from hearing the higher resonances of their voices during front vowel production, the resonances drop in frequency somewhat. Also, there is a tendency for variability to be greater for the F_2 of the less familiar vowels [y] and [ø] than for the familiar vowels [i] and [eᴵ].

Table 2 details the means and standard deviations of the F_3 resonances for the fricatives as imitated by the three speakers under various speaking conditions. For speaker GF the condition involving alginate on the alveolar ridge resulted in lower vocal tract resonances in the F_3 region than for other conditions, but the other two speakers showed little effect. Variability was apt to be higher on unfamiliar syllables and during combined deprivation conditions but not consistently so.

Spectrograms of the imitations of DB's [z] and [X] utterances corresponding to the EMG plots shown in Figure 5 are shown in Figures 7 and 8. Figure 7 (a and b) represents a wide band and a narrow band display of two imitations of "Say zi" as produced under the combined condition of alginate, nerve block, and auditory masking. Figure 8 shows two imitations of [Xi]. The first attempt (Figure 8a) consists of fricative noise, but the formants decline in frequency. It was transcribed as [hʳ] and as [Xʳ] due to its liquid quality. The second attempt (Figure 8b) consists of fricative noise, but voicing continues and it was transcribed as a pharyngeal fricative [ʕ] or as a voiced aspirate. Again, note the difference in superior longitudinal muscle activity in Figure 5.

The spectrograms of the 10 repetitions of the fricative syllables [X] and [ɣ] for each speaker under each condition were also measured. Figure 9 represents the plots for each speaker of F_3 frequencies across 13 trials (10 repetitions and the 3 tokens in the initial list). There is no systematic change that would indicate the presence of learning.

Perceptual Analysis

The purpose of obtaining listener judgments of the speaker imitations was to investigate the perceptual effects of the various speaking conditions on speakers' ability to imitate the familiar and relatively unfamiliar utterances. It was obvious that the familiar syllables were closer to the model under all conditions than were the unfamiliar syllables.

Table 2

F_3 for Fricatives

Based on Three Trials Each

S:DB	ʃ		x		z		ɣ	
	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
N	2567	23	2567	143	2500	41	2242	31
NB	2583	131	2483	165	2683	24	2350	41
NB + M	2450	122	2183	24	2533	24	2133	47
NB, M, & A	2667	47	2200	349	2683	24	2050	147
S:TB								
N	2416	23	2567	47	2267	23	2667	94
NB	2483	24	2400	41	2267	23	2400	108
NB + M	2450	147	2367	62	2333	47	2416	62
NB + A	2467	125	2667	47	2367	47	2550	71
NB, M, & A	2500	0	2267	47	2500	71	2533	201
S:GF								
N	2425	54	2500	204	2333	24	2350	0
M	2375	0	2442	51	2208	59	2342	31
A	2200	35	2133	103	2125	0	2192	42
M & A	2075	54	1867	92	2050	54	2033	47
NB, M, & A	2333	118	2142	23	2257	12	2067	225

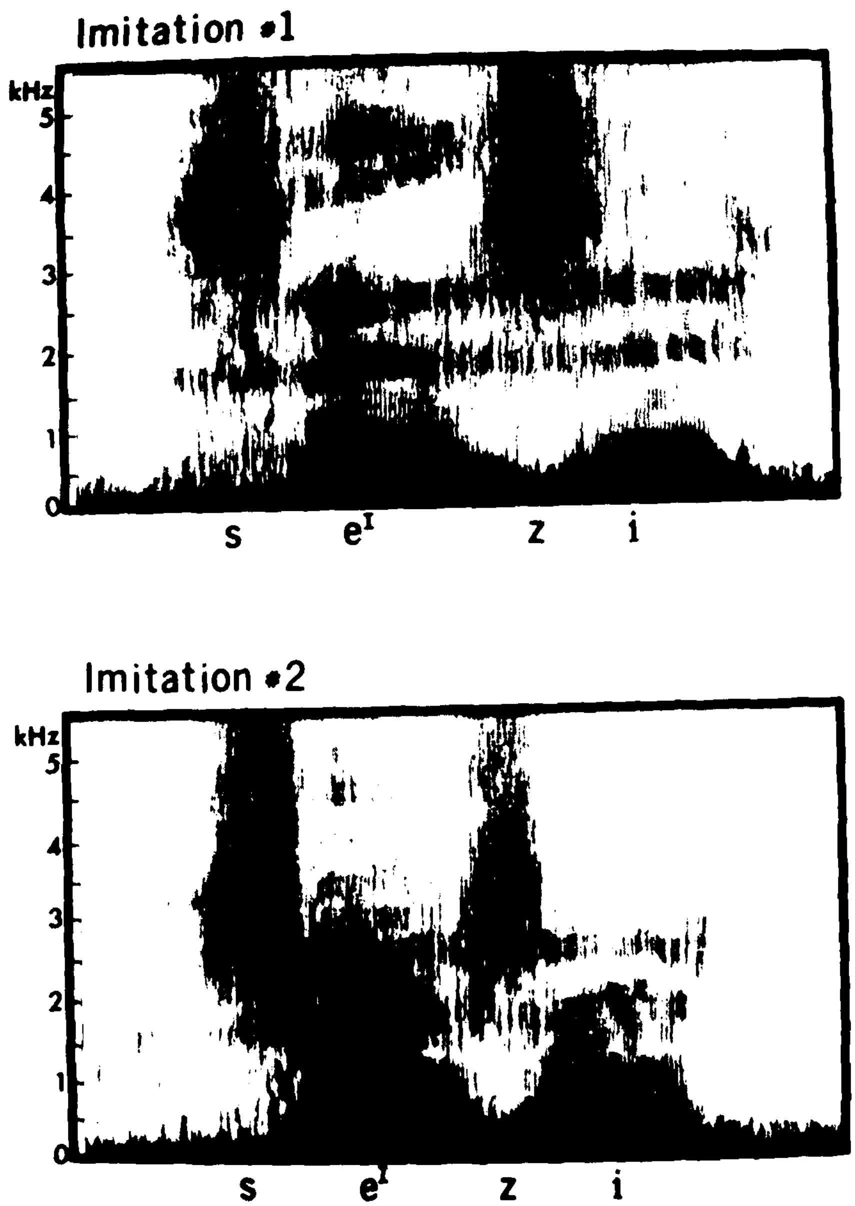


Figure 7a

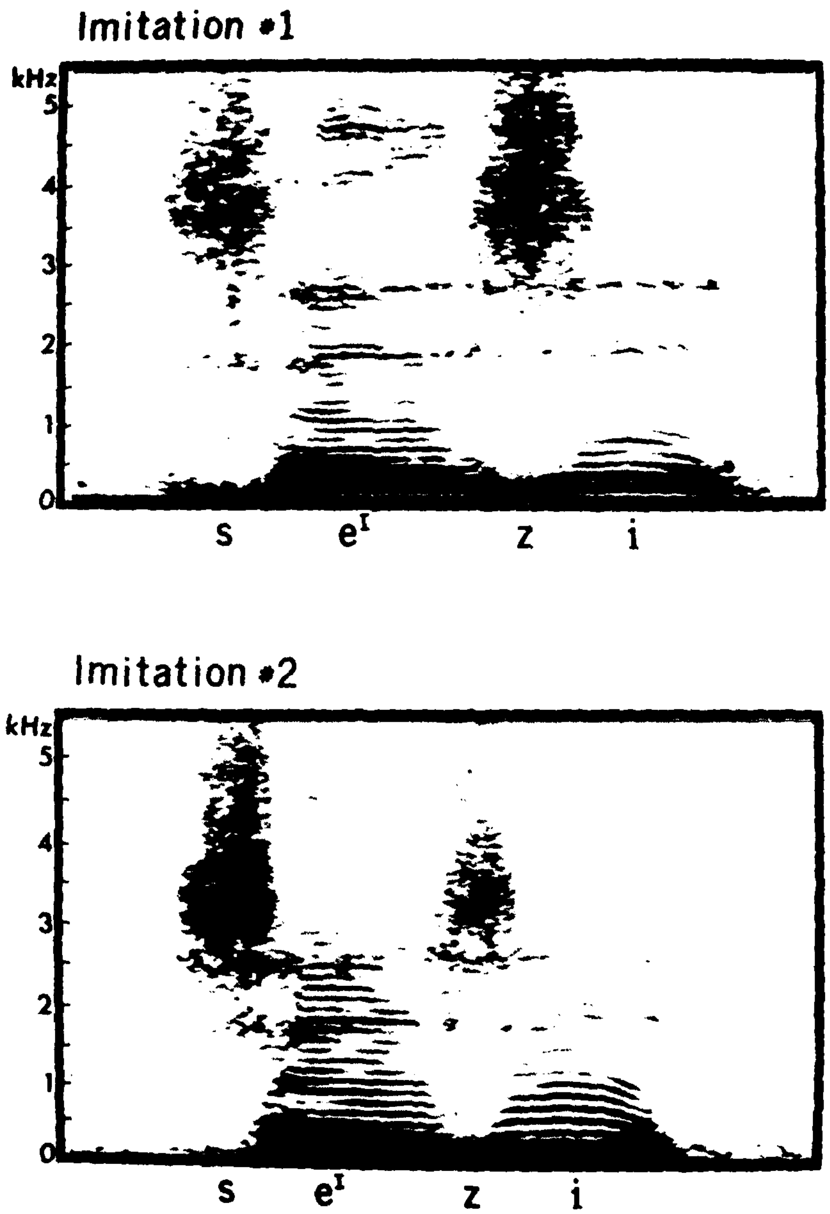


Figure 7b

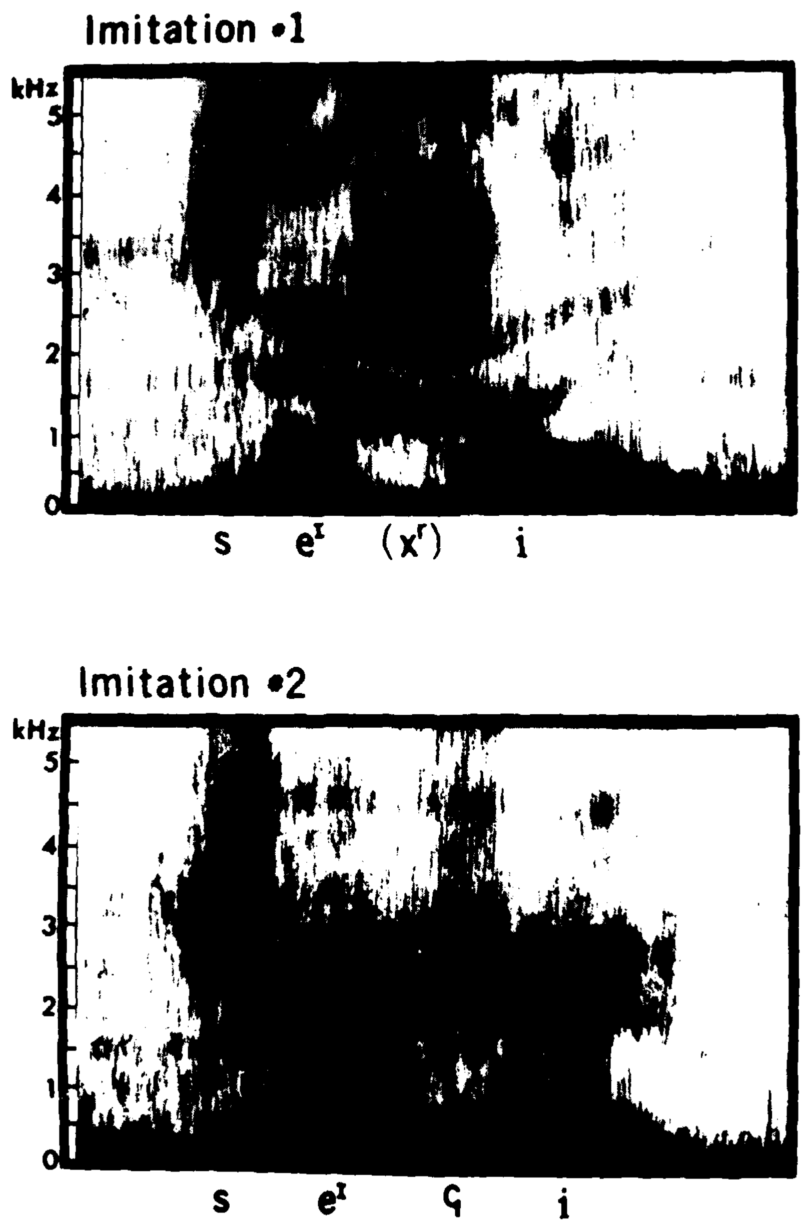


Figure 8a

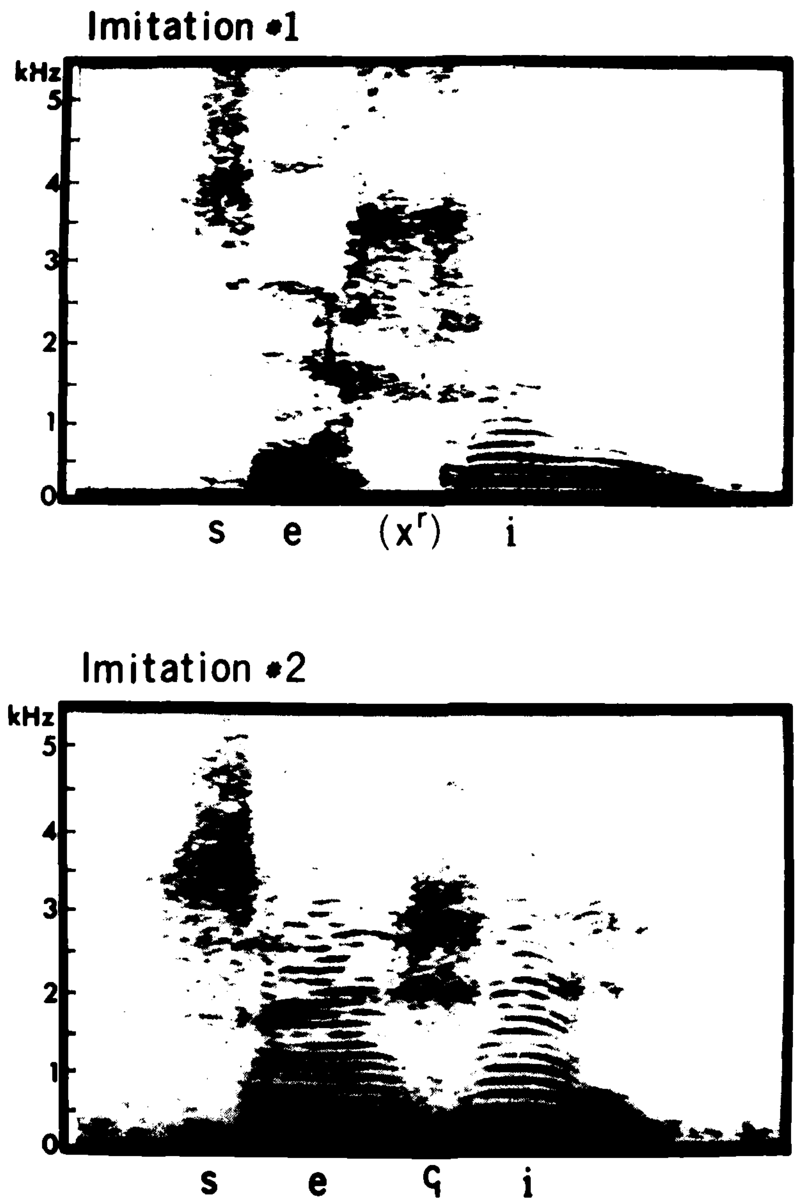


Figure 8b

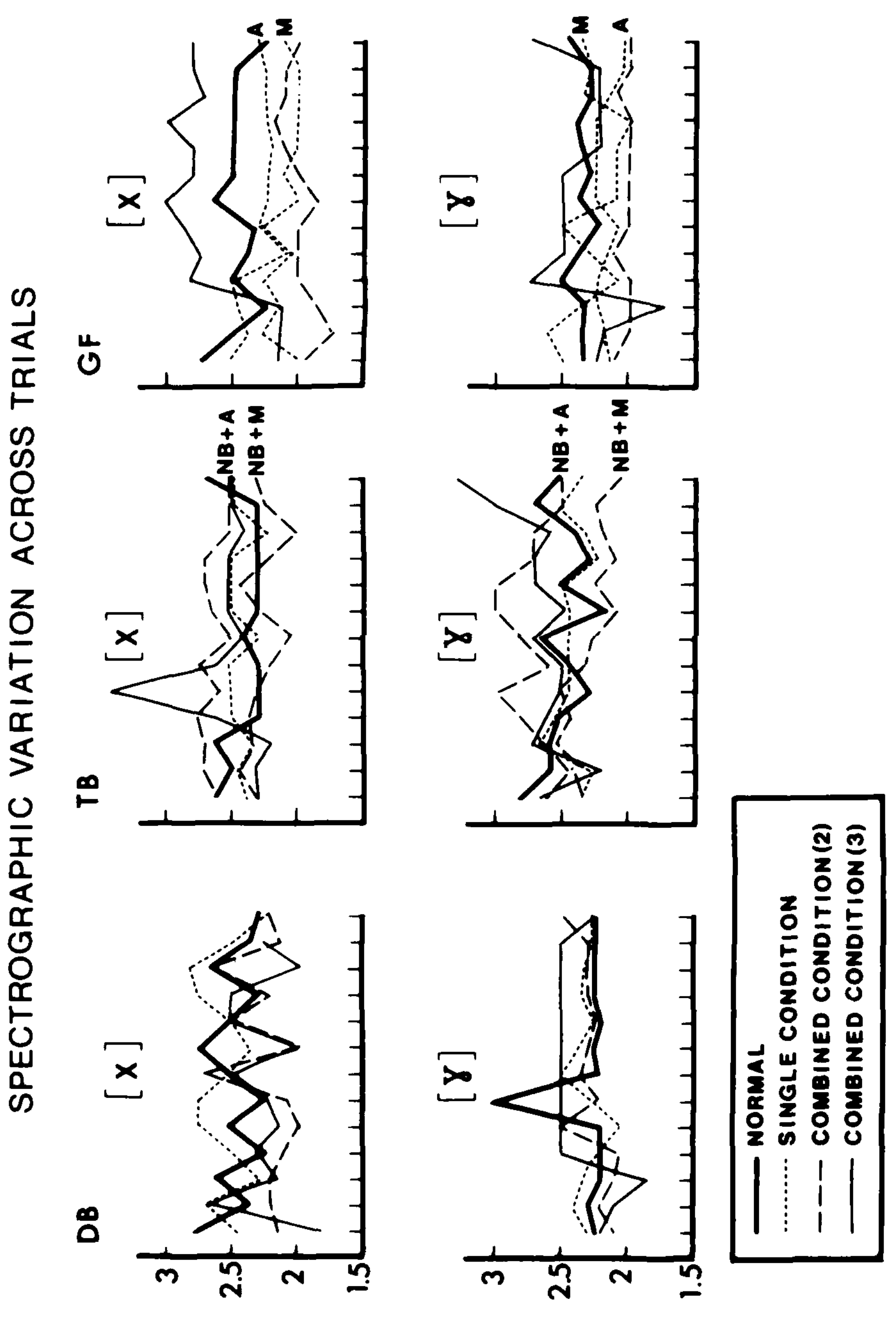

SPECTROGRAPHIC VARIATION ACROSS TRIALS
DB
TB
GF
[x]
[ɣ]
3
2.5
2
1.5
NB+A
NB+M
A
M
NORMAL
SINGLE CONDITION
COMBINED CONDITION(2)
COMBINED CONDITION(3)

Figure 9

Effects of single versus combined sorts of decrement in information. Figure 10 collapses listener judgments of seven speakers--the three speakers with nerve block and the four non nerve-block speakers. This figure represents averaged listener comparisons of altered speaking conditions with the normal condition. Listeners preferred the normal speaking condition to between 60-65 percent of the utterances spoken during any single alteration. The normal condition was preferred to between approximately 80-95 percent of the combined conditions. Thus, decrements in information available to the speaker, although of different sorts, impair speech more in combination than in any single condition.

Effects of speaking condition on familiar versus unfamiliar syllables. To look more closely at how the speaking conditions affect judgments of familiar versus less familiar utterances and judgments of vowel versus fricative syllables, we ran an analysis of variance on all possible paired contrast conditions from the perceptual data for each of the seven speakers. Two within-subject variables were explored: familiarity (familiar versus novel) and syllable type (vowel versus consonant). It can be seen in Table 3 that there is an effect of familiarity for some of the subjects, while there is an effect of syllable type for only one speaker. In some cases there is an interaction of familiarity with syllable type. For three speakers there was no perceptual effect of familiarity, syllable type, or their interaction that reached significance.

In all cases in which there was a significant effect of familiarity, listeners reported stronger perceptual differences between familiar syllables spoken under contrasting conditions than between less familiar syllables spoken under the same contrasting conditions.

In the cases in which there was a significant interaction between familiarity and syllable type, the interaction was speaker-specific. For TB, the novel consonants were perceived to be least changed. This confirms the general perceptual impression that [X] and [ɣ] were rather consistently produced as [ʃ] and [ʒ] despite speaking condition. For the non-nerve-block speaker representing the second order of conditions, S2, the familiar consonants were perceptually more affected than the other syllable types, while for the speaker representing order S4, it was the novel vowels [y-ø] that were perceived to be least changed by conditions.

In summary, differences in imitations of familiar and less familiar vowels and fricatives are more marked according to listener judgments in familiar syllables than unfamiliar ones, and the interaction between familiarity and syllable type depends upon the speaker.

Expert versus student listeners. A possible explanation of the "familiarity" effect is that it is due to the differences in listener familiarity with the sounds, rather than to differences in the productions themselves. In order to examine this possibility, we compared the judgments of an expert listener, naive to the purposes of the study, with the judgments of the student listeners. The expert listener was in general agreement with the naive student listeners in his judgments of relative deterioration of imitations under various speaking conditions. He too found larger differences among the familiar utterances, even though to him all the utterances were more

$$\text{Normal} > \begin{cases} \text{Masking (60\%)} \\ \text{Nerve Block (63\%)} \\ \text{Alginate (65\%)} \end{cases}$$

$$\text{Normal} > \begin{cases} \text{Nerve Block, Masking \& Alginate (81\%)} \\ \text{Masking \& Alginate (82\%)} \\ \text{Nerve Block \& Masking (83\%)} \\ \text{Nerve Block \& Alginate (86\%)} \end{cases}$$

Figure 10

Table 3

Analysis of Variance of Perceptual Data Collected on 7 Speakers
Effects of Familiarity, Syllable Type, and Their Interaction

		FAM.	SYLL.	INT.
Primary Speakers (df 1,9)	DB	F=21.8 p<.001	F=23.2 p<.001	NS
	TB	F=10.6 p<.005	NS	F=5.1 p<.05
	GF	NS	NS	NS
Speakers Without Nerve Block (df 1,5)	S1	NS	NS	NS
	S2	NS	NS	F=7.967 p<.05
	S3	NS	NS	NS
	S4	F=31.506 p<.001	NS	F=11.7 p<.01

familiar than they were to the student listeners. Figure 11 contrasts typical response plots of the expert listener and the student listeners for a familiar and an unfamiliar syllable. For the [ʃ], listener scores reflected a decrease in listener preference as the speaking conditions got more complex, especially upon the addition of auditory masking and alginate, while for [ɣ], scores varied less from one condition to another.

The phonetician who served as the model speaker listened to the lists of 24 utterances spoken under normal conditions, (3 x 8 tokens) by the three primary subjects and the four additional subjects, and judged each imitation to be 1) Americanized or American, 2) Almost Americanized, 3) Neither American nor foreign, 4) Almost foreign, or 5) Authentic Foreign Accent. All speakers were judged to produce ordinary American productions of all familiar syllable types.

Table 4 summarizes for all seven speakers the percentages of familiar and less familiar utterances judged by the phonetician to be correct. The familiar utterances are counted as correct if they are judged to fall within the American English system. The less familiar utterances are counted as correct if they are judged to be within the foreign sound system.

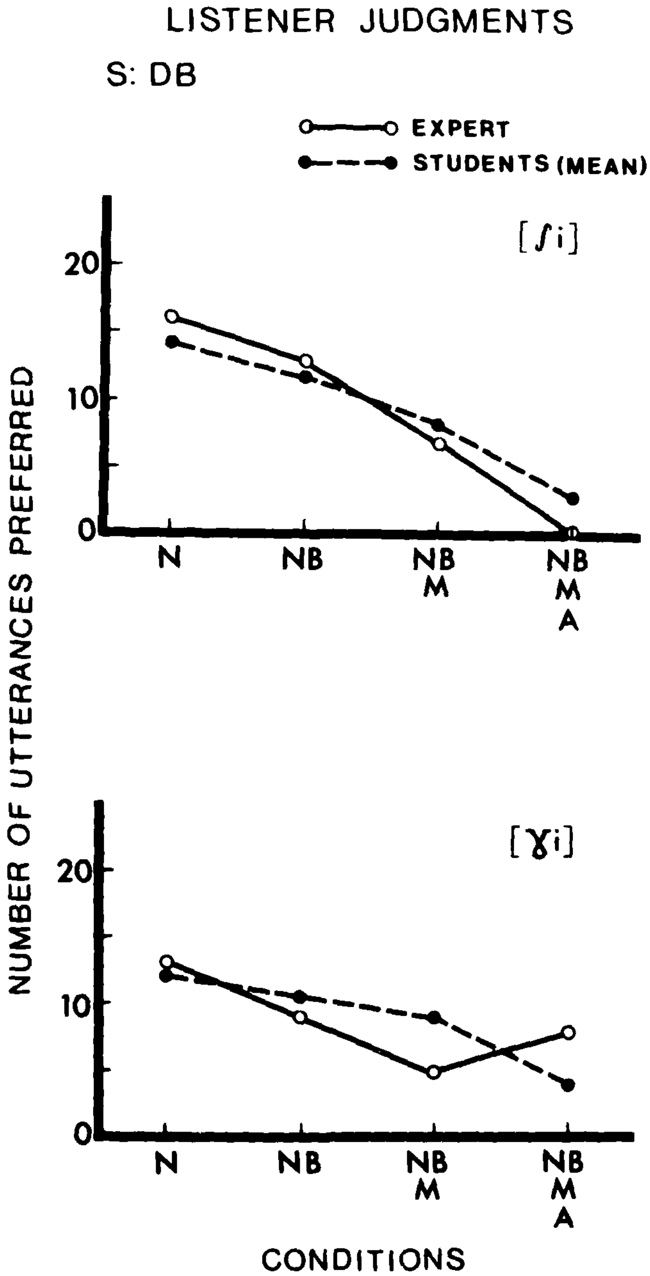
Table 4

Judgments Made by Phonetician of Utterances Spoken Under Normal Speaking Conditions by 7 Speakers

Familiar Utterances Judged Within American English System (Correct)			Less Familiar Utterances Judged Within Foreign Sound System (Correct)		
1.	[ʃ]	100%	1.	[ɸ]	33%*
2.	[e^{I}]	90%	2.	[χ]	29%+
3.	[i]	81%	3.	[ɣ]	5%+
4.	[z]	75%	4.	[y]	0%*

N = 21
7 speakers x 3 tokens

*Never Americanized; 57% Judged Almost Foreign
+Americanized; 14% [χ] and 29% [ɣ]



LISTENER JUDGMENTS
S: DB
EXPERT
STUDENTS (MEAN)
[ʃi]
[ɣi]
NUMBER OF UTTERANCES PREFERRED
20
10
0
N
NB
NB M
NB M A
CONDITIONS

Figure 11

Even under normal speaking conditions, [ʃ] was judged to be acceptable despite acoustic variation, while [z] was more vulnerable to perceptual inconstancy. American students were apt to Americanize the novel fricatives, but never Americanized the rounded front vowels, which were judged more than half the time to be almost foreign or correct in production.

Effects of conditions on syllable types. The nerve block condition results in a rather small effect across syllables (Figure 12). A 50% preference would be expected by chance. Listeners judged the familiar utterances of DB to be a bit better under normal conditions. For TB the perceptual difference is increased and includes the less familiar vowels. GF produced no perceptually different imitations under nerve block except for the syllables [pi] and [pe^{I}]. However, we have no nerve block alone condition for GF, so with the nerve block, the subject could not hear himself and his palate was thickened with alginate. Under this combined condition, the addition of a nerve block was noticeable on 72% of the [i] and [e^{I}] utterances.

Auditory masking, too, has only a small effect on listener judgments and affects different syllables for different speakers: DB the fricatives, TB the vowels, and GF all but the novel vowels. For the tapes of the non-nerve block speakers, we asked listeners to check the worse imitation but to mark it with an X if it were much worse. Imitations under auditory masking were judged to be much worse in [X] for speaker #1, for [ɣ] for speakers #2 and #4, but not much worse for any syllable of speaker #3.

Alginate placed on the alveolar ridge is more disruptive, according to perceptual judgments, than is either auditory masking or lingual nerve block. The altered vocal tract produces a more noticeable decrement, on the average, to the fricatives than to the vowels. It is especially disruptive to [ʃ] but also affects [i], while sparing (perhaps even facilitating) [ø] and [y].

Summary of Results

Putting the results of the various analyses together, we find that:

(1) For four speakers, familiar syllables were more noticeably affected by adverse feedback conditions than were less familiar syllables, but for three out of seven speakers, familiarity of syllables did not contribute to perceived differences among conditions. When there was an effect of familiarity, the familiar fricatives were more affected than vowels by adverse speaking conditions;

(2) Speakers made perceptually intelligible imitations of familiar syllables under all speaking conditions, although more acoustic variation was evident for [ʃ] than for the other English phones;

(3) Speakers differed in their ability to imitate non-familiar syllables with two subjects (DB and GF) making a variety of attempts to produce [X] and one subject (TB) consistently substituting [ʃ] for [X] and [ʒ] for [ɣ];

(4) Non-English [y] and [ø] were more closely approximated than were the fricatives by all three speakers, with listeners perceiving the produc-

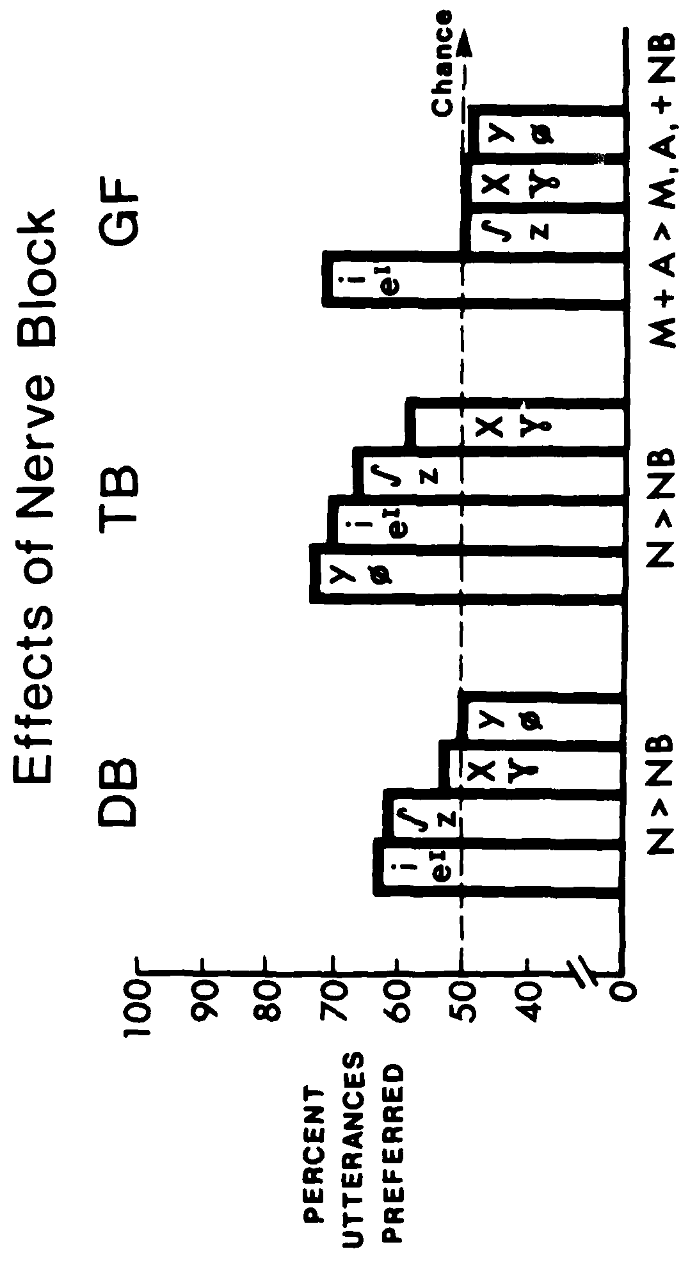

Effects of Nerve Block
DB
TB
GF
PERCENT
UTTERANCES
PREFERRED
100
90
80
70
60
50
40
0
Chance
N > NB
N > NB
M + A > M, A, + NB

Figure 12

tions as similar under all speaking conditions. Appropriate OO activity was evident for rounded vowels and the F_2 was usually lower than for their unrounded cognates as would be expected given the longer vocal tract resulting from lip rounding. In agreement with these indications of lip rounding, the less familiar vowels were judged by the phonetician to be almost foreign under normal speaking conditions;

(5) Listeners, whether expert or naive, preferred the imitations made under normal speaking conditions to those under any of the adverse speaking conditions. Imitations produced with lingual anesthesia were preferred to those produced with masking or with Alginate. Combined conditions were judged worse than any single condition;

(6) Nerve block produced a small effect on all syllable types;

(7) Auditory masking affected some syllable types more than others, depending on the speaker. In general, the F_2 frequency for vowels was lower with masking;

(8) Vocal tract shape change affected [ʃ] and [i] particularly, with both spectrographic and EMG evidence of tongue retraction;

(9) There was no evidence of learning. Subjects apparently knew how to approximate [y] and [ø] without trials, but for [X] and [ɣ] they tried and failed in the time given (13 trials).

DISCUSSION

Studies of speaker compensation under difficult speaking conditions have concurred in their results, indicating that speakers are able to produce acceptable speech patterns despite bite blocks between the teeth (Lindblom, Lubker, & Gay, 1979; Fowler & Turvey, 1980) forcing a change in motor activity and despite conditions such as auditory masking and anesthesia, changing sensory information (Borden, 1979). The present study is the first, however, to manipulate the familiarity of the phonetic material. By contrasting familiar utterances with less familiar utterances, the importance of learning to motor control may be evaluated since the familiar utterances have been well practiced relative to the less familiar utterances.

The problem lies in disambiguating the perceptual and productive aspects of the control systems. To produce a skilled motor act such as a well-learned speech event, the speaker presumably makes reference to an internal representation of the utterance in the form of a perceptual target and then effects the appropriate motor coordinations according to known production rules. To produce a relatively unfamiliar motor act such as a foreign speech event, the speaker presumably refers to a more poorly formed perceptual target and enacts a motor program based on less well known production rules. Familiarity, then, influences both the perceptual target and the production.

The finding in the present study, that more familiar utterances are in some subjects more vulnerable to changes in speaking conditions than are the

less familiar utterances, lends support to the hypothesis that criteria for the perceptual targets that speakers have internalized are more detailed for familiar than for unfamiliar utterances, and that the motor programs used to produce them are more refined and well practiced. Loss of information needed to sharpen the match between speakers' actual output and intended output may result in small acoustic differences perceptible by listeners. The same loss of information about production of less familiar utterances resulted in these speakers producing less perceptually noticeable effects across speaking conditions. It follows from the same hypothesis that the less familiar utterances would be represented in the speakers' mind with a set of auditory, oro-sensory criteria that might be less well defined than for familiar utterances, as well as with a more poorly practiced set of production strategies. The use of oro-sensory information for the fine shaping of speech events has been suggested in the work of Stevens and Perkell (1977).

How is one to infer that these differences arise from differences among the productions of the speakers and not simply from the expectations of the listeners making the perceptual judgments? We know that listeners tend to categorize allophonic variations according to the phoneme systems of their own language and, further, tend to ignore small acoustic differences within their own phonemic categories (Liberman, Harris, Hoffman, & Griffith, 1957). According to the principles of categorical perception, then, English-speaking listeners would be expected to ignore differences among English-like utterances that they might notice among more foreign utterances. To the degree that the listeners in this study noticed the differences in the familiar utterances more than the unfamiliar, it can be inferred that the differences were real: they existed in the speech productions and not solely in the perceptions made by listeners.

Further support for this inference comes from the agreement with English listener judgments of an expert listener for whom the "unfamiliar" utterances were native to his language. Finally, the unfamiliar syllables (especially the consonants) were so poorly imitated and so variable, even under normal speaking conditions, that the differences between speaking conditions were relatively unimportant, whereas the imitations of the familiar syllables were remarkably good in all conditions, but small differences across speaking conditions were perceptible.

Familiarity of the phonetic material was not a significant factor for some of the speakers, indicating that for these speakers, loss of information about their own speech made no more difference in their performance, whether the performance involved well-learned or novel speech productions. The implication here is that control was essentially preplanned; with little evidence of the fine tuning of the well-learned utterances shown by three of the seven speakers.

The idea of feedforward or preprogrammed control of motor systems in speech is consistent with recent findings in the motor control literature cited in the Introduction. The compensatory motor patterns evidenced by people and by animals despite conditions that require changes in motor coordination or that remove sensory information argue for a control system with extremely rapid adaptability features. Some theorists account for the compensatory power of such motor systems by suggesting that under difficult

circumstances the motor plan is compared to the motor performance through afferent systems (Evarts, 1971; MacNeilage, 1970) or that the efferent program is simply simulated and matched with simulated afferent information, thus enabling the system to adjust by prediction without waiting for actual performance (Lindblom et al., 1979). Other theorists account for the compensations by suggesting that the equilibrium points for final positions are specified and any interference with one part of the system is adjusted for by another part of what is essentially a vibratory system (Bernstein, 1967; Kelso & Holt, 1980) or a coordinative structure (Fowler & Turvey, 1978).

There was no evidence of learning in the 13 trials produced by each subject of the less familiar fricatives [X] and [γ]. The speakers apparently failed to make use of information provided through feedback mechanisms to quickly shape a novel speech gesture. The well-programmed production rules appropriate to the speakers' language seemed in many cases to override any attempts to match a new perceptual target. It is impossible to determine from this study whether the difficulty in making the appropriate changes arises from a poor perceptual image of the target, from inadequate and poorly practiced production rules, or from a combination of perception and production factors.

There was less difficulty with [py] and [pϕ] according to listener judgments. The internal formation of some auditory-perceptual target may well have been less demanding than for the unfamiliar fricatives. If the perceptual image is easier to elicit, might it be because the production rules are not far from sounds produced in English? Although there might be a slight difference in tongue elevation and fronting for [i] and [y], the gesture itself is not novel, nor is the gesture of lip rounding. Subjects seemed to make generally acceptable [py] and [pϕ] and the fact that they were so little affected by loss of auditory feedback indicates that the strategy taken was relatively simple (lip rounding; none of the conditions affected the lips) and may have been controlled by feedforward or open loop instructions to round the lips.

The implications of this kind of study for second language learning are obvious. We know a bit more about the ways in which perception precedes production in children learning their first language (Menyuk & Anderson, 1969; Strange & Broen, 1980; McReynolds, 1978) than we do about the ways in which perception and production may interact in adults learning a new language (MacKay, 1970; Goto, 1971; Williams, 1974; Borden, 1980).

These data suggest that for adults the basic articulation responses for speech may operate under an automatic open loop motor system, with the fine tuning of such responses resting upon the availability of a well-defined perceptual target and information on the sounds and oral sensations produced--at least for the production of continuants such as vowels and fricatives.

Future research might explore whether such feedback information can contribute to more rapid speech events as well as continuants. Also, it would be interesting to try to measure separately the development of a new perceptual target or image and the development of new motor strategies, in order to evaluate their respective contributions to the production of new speech sounds.

REFERENCES

Bernstein, N. A. The coordination and regulation of movements. London: Pergamon Press, 1967.

Borden, G. J. An interpretation of research on feedback interruption in speech. Brain and Language, 1979, 7, 307-319.

Borden, G. J. Use of feedback in established and developing speech. In N. J. Lass (Ed.), Speech and language: Advances in basic research (Vol. III). New York: Academic Press, 1980, 223-242.

Borden, G. J., & Gay, T. Temporal aspects of articulatory movements for s-stop clusters. Phonetica, 1979, 36, 21-31.

Borden, G. J., Harris, K. S., & Oliver, W. Oral feedback I. Variability of the effect of nerve-block anesthesia upon speech. Journal of Phonetics, 1973, 1, 289-295.

Chase, R. A. Abnormalities in motor control secondary to congenital sensory defects. In J. F. Bosma (Ed.), Symposium on oral sensation and perception. Springfield, Ill.: Charles C. Thomas, 1967

Evarts, E. V. Feedback and corollary discharge: A merging of the concepts. Neurosciences Research Program Bulletin, 1971, 9, 86-112.

Fowler, C. A., & Turvey, M. T. Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978.

Fowler, C. A., & Turvey, M. T. Immediate compensation in bite-block speech. Phonetica, 1980, 37, 306-326.

Goodwin, G. M., & Luschei, E. S. Effects of destroying the spindle afferents from the jaw muscles upon mastication in monkeys. Journal of Neurophysiology, 1974, 37, 967-981.

Goto, H. Auditory perception by normal Japanese adults of the sounds "L" and "R." Neuropsychologia, 1971, 9, 317-323.

Hamlet, S. L., & Stone, M. Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. Journal of Phonetics, 1976, 4, 199-218.

Harris, K. S., & McGarr, N. S. Relationships between speech perception and speech production in normal hearing and hearing impaired subjects: Implications for training speech of the hearing impaired. In J. D. Subtelny (Ed.), Speech assessment and speech improvement for the hearing impaired. Washington, D.C.: A. G. Bell Association, 1980, 316-337.

Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and techniques. Haskins Laboratories Status Report on Speech Research, 1971, SR-25/26, 73-86.

Kelso, J. A. S., & Holt, K. G. Exploring a vibratory systems account of human movement production. Journal of Neurophysiology, 1980, 43, 1183-1196.

Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research, 1973, SR-33, 173-183.

Lane, H. L., & Tranel, B. The Lombard sign and the role of hearing in speech. Journal of Speech and Hearing Research, 1971, 14, 677-709.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 54, 358-368.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics, 1979, 7, 147-161.

Lindblom, B., & Sundberg, J. Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of Acoustical Society of America, 1971, 50, 1166-1179.

Locke, J. L. Oral perception and articulation learning. Perceptual and Motor Skills, 1968, 26, 1259-1264.

MacKay, D. G. How does language familiarity influence stuttering under delayed auditory feedback? Perceptual and Motor Skills, 1970, 30, 655-669.

MacNeilage, P. The motor control of serial ordering in speech. Psychological Review, 1970, 77, 182-196.

McReynolds, L. V. Behavioral and linguistic considerations in children's speech production. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic. Cambridge, Mass.: MIT Press, 1978, 127-164.

Menyuk, P., & Anderson, S. Children's identification and reproduction of /w/, /r/, and /l/. Journal of Speech and Hearing Research, 1969, 12, 39-52.

Osberger, M. J., & Levitt, H. The effect of timing errors on the intelligibility of deaf children's speech. Journal of the Acoustical Society of America, 1979, 66, 1316-1324.

Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. Science, 1978, 201, 1235-1237.

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. Frequency analysis of Dutch vowels from 50 male speakers. Journal of the Acoustical Society of America, 1973, 53, 1093-1101.

Raphael, L. J., Bell-Berti, F., Collier, R., & Baer, T. Tongue position in rounded and unrounded front vowel pairs. Language and Speech, 1979, 22, 37-48.

Ringel, R. L., & Steer, M. Some effects of tactile and auditory alterations on speech output. Journal of Speech and Hearing Research, 1963, 6, 369-378.

Siegel, G. M., & Pick, H. L. Jr. Auditory feedback in the regulation of the voice. Journal of the Acoustical Society of America, 1974, 56, 1618-1624.

Stevens, K. N., & Perkell, J. S. Speech physiology and phonetic features. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977, 323-345.

Strange, W., & Broen, P. A. Perception and production of approximant consonants by three-year-olds: A first study. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 2) Perception. New York: Academic Press, 1980, 117-154.

Taub, E., & Berman, A. J. Movement and learning in the absence of sensory feedback. In S. J. Freedman (Ed.), The neuropsychology of spatially oriented behavior. Homewood, Ill.: Dorsey Press, 1968, 173-192.

Williams, L. Speech perception and production as a function of exposure to a second language. Unpublished doctoral dissertation, Harvard University, 1974.

ORTHOGRAPHIC VARIATIONS AND VISUAL INFORMATION PROCESSING*

Daisy L. Hung+ and Ovid J. L. Tzeng+

Abstract. Based upon an analysis of how graphemic symbols are mapped onto spoken languages, three distinctive writing systems with three different relations between script and speech relationships are identified. They are logography, syllabary, and alphabet, developed sequentially in the history of mankind. It is noted that this trend of development seems to coincide with the trend of cognitive development of children. This coincidence may imply that different cognitive processes are required for achieving reading proficiency in different writing systems. The studies reviewed include experiments on visual scanning, visual lateralization, perceptual demands, word recognition, speech recoding, and sentence comprehension. Results from such comparisons of reading behaviors across different orthographies suggest that human visual information processing is indeed affected by orthographic variation, but only at the lower levels (data-driven, or bottom-up processes). With respect to the higher-level processing (concept-driven, or top-down processes), reading behavior seems to be immune to orthographic variations. Further analyses of segmentation in script as well as in speech reveal that every orthography transcribes sentences at the level of words and that the transcription is achieved in a morphemic way.

INTRODUCTION

Ever since Rozin, Poritsky, and Sotsky (1971) successfully taught a group of second-grade nonreaders in Philadelphia to read Chinese, the question has been repeatedly raised: If Johnny can't read, does that mean Johnny really can't read in general or Johnny just can't read English in particular? To the

---

*A revised version is to appear in Psychological Bulletin.

+Also University of California, Riverside.

Acknowledgement: The writing of this paper was supported in part by a grant from the National Institute of Education (NIE-G-81-0055) and in part by a research grant from the Academic Senate of the University of California, Riverside, to the second author. The final form of this paper has benefited tremendously from discussions with Alvin Liberman, Ignatius Mattingly, Don Shankweiler, and other members of Haskins Laboratories. The authors would like to thank Bruno Repp for his critical reading of an earlier version of this manuscript. We should also acknowledge with great appreciation the help of Katherine Miller, whose excellent editorial skills made this paper much more readable.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

reading specialists, educational psychologists, and cognitive psychologists who are interested in the visual information processing of printed materials such a question is of practical as well as theoretical importance with respect to the understanding of reading behavior. At the practical level, is it true that some writing systems are easier to learn than others, and to what degree can dyslexia be avoided given that a certain type of writing system happens to be used for a certain type of spoken language? At the theoretical level, one must start to untangle the relations between script and speech by uncovering strategic differences at various levels of information processing (feature extraction, letter identification, word recognition, etc.) in the reading of different writing systems. These analyses may result in a new form of linguistic determinism (cf. Scribner & Cole, 1978; Tzeng & Hung, 1980).

It is conceivable that reading different scripts entails different processing strategies. Paivio (1971) has gathered much evidence that meanings of words and of pictures are retrieved via different routes. Thus, one may speculate that, depending on how spoken languages are represented by printed symbols, readers have to develop different processing strategies in order to achieve proficiency in reading. Failure to develop these strategies may result in a certain type of dyslexia that may be avoided when learning to read another script. For example, because of the close grapheme-sound relation, alphabetic script may require beginning readers to pay special attention to phonetic structure. Children who have not developed the appropriate "linguistic awareness" (Mattingly, 1972) of such a phonetic structure may become nonreaders. The same children, who are classified as dyslexic under an alphabetic system, may encounter no problem in learning to read a sign script such as Chinese logographs.

The idea of teaching the dyslexic to read Chinese is by no means new. According to Hinshelwood (1917), Bishop Harmon suggested that the ideal therapy for this disorder was to teach dyslexic children Chinese characters, because Chinese is a sign script where each word was its own symbol. The success of Rozin et al. (1971), though it has not gone uncriticized (Ferguson, 1975; Tzeng & Hung, 1980), undoubtedly reinforces this idea and seems to point to the possibility that dyslexia may not characterize visual-verbal association in general. Hence, for a general understanding of reading behavior, cross-language comparisons of visual information processing strategies should provide valuable clues to the underlying mechanisms and processes involved in reading.

We will critically review some recent studies pertinent to the issue of comparative reading. We will begin by discussing different orthographic rules for mapping written scripts onto speech in various languages. Then we will examine results of experiments that were conducted to find out whether these orthographic variations have any effect on visual information processing. Finally, we will draw some tenable conclusions about the relations between orthography and reading.

RELATIONS BETWEEN SCRIPT AND SPEECH

The relation between written scripts and spoken languages seems so close that one would expect that anyone who is able to speak should be able to read.

But this is simply not the case. For all normal children, learning spoken language seems to require no special effort. From the time the child is able to emit his first sound, he is tuned to engage actively in the language acquisition game, and the process seems to be spontaneous. Some psycholinguists (e.g., McNeill, 1970) even suggest that the language acquisition device is prewired biologically in our genetic program and that the language environment serves as a stimulus releaser to allow this program to unfold. Learning to read, on the contrary, requires a relatively long period of special training and depends heavily on intelligence, motivation, and other social-cultural factors. And even with so much effort directed toward the acquisition of reading skills, not every child is blessed with the ability to read.

There is a general consensus that written languages evolve much later than spoken languages and that in some way the former are attempts to record the latter. Increasingly complicated and sophisticated living experience renders oral communication an unsatisfactory mediator for cultural and social transmission. If one is able to transcribe spoken language visually into some kind of graphic representation, then communication can overcome the limitations of space and time that are usually imposed on the spoken sound. Since there are many levels of representation of spoken language, the transcription of spoken language into visual symbols can be achieved in many different ways. If we look back at the history of mankind, we soon discover that the evolution of writing systems proceeds in a certain direction. In a sense, the transcription starts at the deepest level, the conceptual gist, and gradually shifts outward to the surface level, the sounds. At each step, unique and concrete ways of representing meaning give way to a smaller but more general set of written symbols. In other words, writing efficiency is achieved by sacrificing the more direct link to the underlying meaning; consequently, the grapheme-meaning relation becomes more abstract.

Primitive men wrote (or more precisely, carved) on rocks, tortoise shell, cave walls, and so on, to achieve some form of communication. These drawings were usually pictures of objects that immediately evoked meaningful interpretations. A general idea (*sememe*), rather than a sequence of words in a sentence, was expressed via object drawing. Thus, *semasiography writes* concepts directly without the mediation of spoken language. Archaeologists have discovered these rock paintings and carved inscriptions in many parts of the world (Asia, Europe, Africa, America, Australia Oceania). From them they are able to reconstruct and speculate about the life styles of these early men (Gelb, 1952). However, picture drawing as a communication tool has many obvious difficulties. First of all, not everyone is capable of good drawing. Second, it is difficult to draw pictures that express abstract concepts. Third, different ways of arranging objects within a picture result in different interpretations. Finally, an unambiguous picture (e.g., a map telling the location of food resources) can be disadvantageous. Thus, new systems had to be invented.

The next step is important and insightful and should be regarded as one of the most important achievements in the history of mankind. Instead of expressing a general idea by drawing a picture, symbols were then invented to represent the spoken language directly. First, there were pictograms, (e.g., 木 for tree), which were carried over from the previous stage of

picture drawing. Then, there were ideograms, which are frequently formed by putting several pictograms together to suggest an idea: for instance, putting two trees together side by side to mean GROVE (林) and stacking three trees together to mean FOREST (森). Thus, by the principle of metonymy, many ideograms were invented to represent ideas and feelings of various kinds.[1]

But even with this new invention, there were still difficulties in forming characters to represent abstract concepts. This need led to the invention of phonograms, which were typically made up of two or more components, one of which was used not for its semantic content but for its phonetic value. The reader gets a hint as to the character's meaning from the semantic component (called the signific) and to its sound from the phonetic component. With these three methods and the combination of them, a large number of characters may be created to represent all words used in the spoken language. This is exactly how the Chinese logographic system was formed (Wang, 1973). Some examples of the formation of pictograms and of phonograms in Chinese are illustrated in Figure 1. Similar principles were also used in ancient Egyptian hieroglyphics and hieratics (Gelb, 1952). For example, the cartouche (an oval or oblong figure) was used as a signific to enclose the syllabic spelling of a monarch's name. It should be noted at this point that once the concept of sound writing was conceived and appreciated, it immediately became a powerful tool for inventing new characters; it was so powerful that nowadays a majority of Chinese characters are phonograms (Wang, in press).

Chinese logographs actually map onto spoken language at the morphemic level. Such a one-to-one grapheme-morpheme relation in the logographic system requires that there must be distinctive characters corresponding to each morpheme. The inevitable consequence is that one has to memorize thousands of these distinctive characters before one is able to read. Furthermore, writing is tedious and slow. Printing and typing demand too much effort and time, and in an era of mechanization and computerization cries for change are echoed at every level of the Chinese scientific community. This is not the place to enter the debate for or against the character reform currently taking place in the People's Republic of China. Suffice it to say that the logographic script, with so close a grapheme-meaning relation, has its difficulties and is under a great deal of technological pressure. However, one should bear in mind that this does not mean that logographic scripts are in any sense less advanced than alphabetic scripts. Evolutionary fitness should be defined in terms of the particular environment. The intrinsic virtue of Chinese logographs cannot be outweighed by technological difficulties that may easily be overcome by further technological advancements. What we need to find out is how the logographic scripts affect reading behaviors.

We have already noted the power of representing sound. It takes only a small step to go from the rebus[2] system to the syllabic system, in which every written symbol denotes a syllable in the spoken language. As we can see from cuneiform syllabaries, west Semitic syllabaries, Aegean syllabaries, and Japanese syllabaries, the design feature is a close symbol-sound relation. Thus, with a relatively small set of syllable-based symbols one can transcribe an infinite number of spoken sentences. An economy of writing is accomplished and the unit of written language coincides with that of the spoken language. However, there immediately arises the problem of homophones, which are indeed

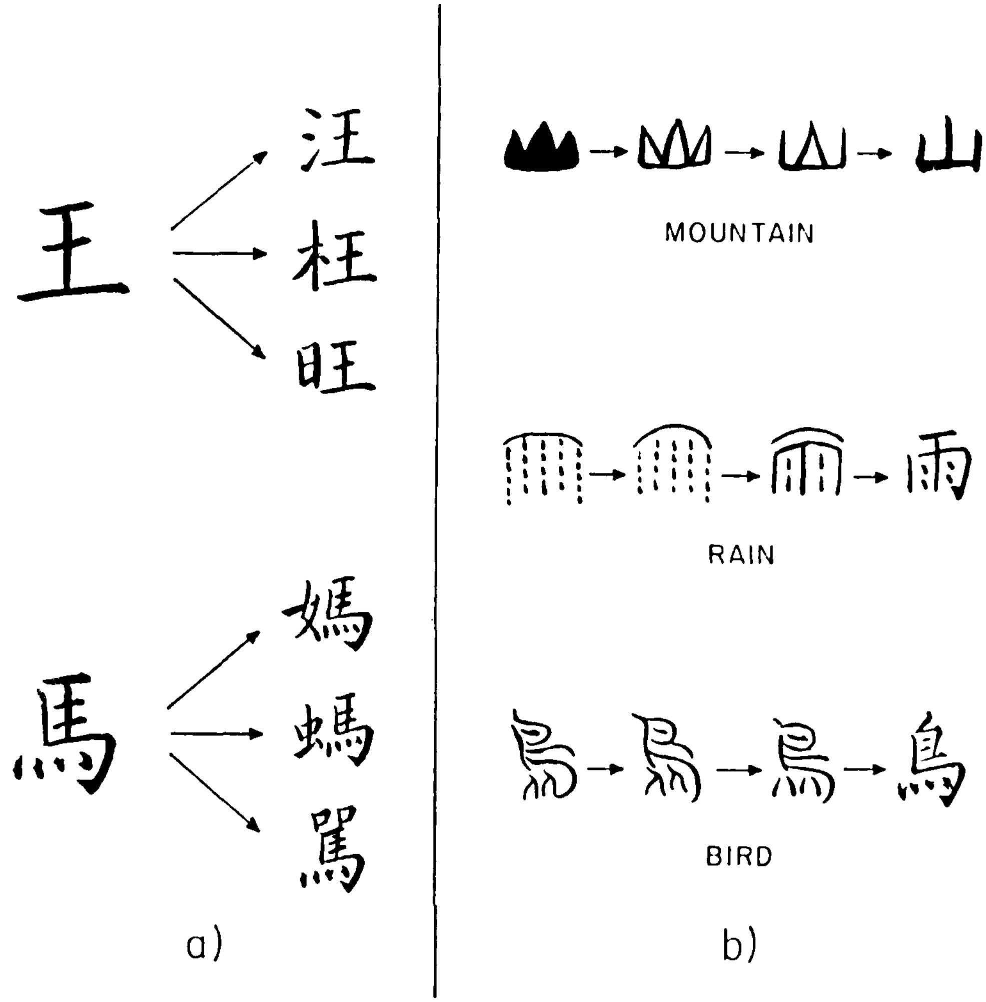


Figure 1. (a) Examples of Chinese phonograms. In the upper panel, the character on the left-hand side is the base character and is pronounced as /wang/ (meaning KING). The three characters on the right are derivatives that contain the base character as a clue to their pronunciations. In fact, they are pronounced as /wáng/, /wǎng/, and /wàng/ from top to bottom, meaning THE BARKING SOUND OF DOGS (or alternatively, DEEP AND WIDE), NOT STRAIGHT, and PROSPERITY for the three characters, respectively. In the lower panel, the base character on the left is pronounced as /má/. It means HORSE, and it is a pictogram by itself (see Wang, 1973). Similarly, the three derivative characters on the right are pronounced as /mā/, /mǎ/, and /mà/, meaning MOTHER, ANT, and TO SCOLD, respectively. Thus, if a reader knows how to pronounce the base characters, he can guess at the pronunciations of the derived phonograms that contain the base character as a partial component. However, one should be cautious in making generalizations because in many cases the base character only gives a clue to the sound of a particular phonogram (sometimes the clue refers only to the vowel ending) and the tonal patterns (ˉ, ˊ, ˇ, ˋ) are not included. (b) Examples of pictograms and their transformation through hundreds of years.

a nuisance even with the contextual cues provided in reading (Suzuki, 1963). This problem is best exemplified by the Japanese writing system, in which three different types of scripts, namely, kanji, katakana, and hiragana (four if we also count the Roman letters used in many modern Japanese texts, i.e., romaji), are concurrently used in order to overcome the difficulty of homophones. In the Japanese syllabary, the problem was resolved by retaining Chinese logographs, generally referred to as kanji, to be used as the content words. The kana script is a set of symbols representing the syllable sounds of the spoken language; thus, in principle, it can be used to write any word in the Japanese language. The kana script is subdivided into two types, hiragana script and katakana script. The former, a more cursive style, is the script used for writing the grammatical particles and function words; the latter is mainly used to write loan words (foreign words such as television). These three different scripts, kanji, hiragana, and katakana, are used concurrently in a text. Because they have different writing styles and serve different linguistic purposes, reading is probably facilitated by these distinctive visual cues. On the other hand, all the difficulties associated with the logographic script arise once again. It is no wonder that over the last 30 years, the Japanese government has been making every effort to eliminate Chinese characters in their writing system. However, the close grapheme-morpheme relation represented in a Chinese character has enough intrinsic value in facilitating visual reading that these attempts to abandon the Chinese characters have not been successful. Ironically, instead of reducing the number of characters, the Ministry of Education was recently forced to add five more characters to their allowable list.

For most of the Indo-European languages, the writing system patterned after the Greek system, and further evolved to an alphabetic system, with the number of written symbols further reduced. A full alphabet, marking vowel as well as consonant phonemes, developed over a period of about 200 years during the first millenium B.C. in Greece (Kroeber, 1948). The transition from the syllabic to the alphabetic system marks another gigantic jump with respect to the script-speech relation. The discovery of vowel letters, which form the basis of the analytical principle of an alphabetic system, has been characterized as something of an accident rather than a conscious insight (Gleitman & Rozin, 1977). As a sound-writing script, an alphabetic system maps onto speech at the level of the phoneme, a linguistic unit smaller than the syllable but larger than an articulatory feature. The problem of homophones was solved in some languages (e.g., English) by simultaneously taking into account the lexical root of each word. The consequence is that the grapheme-sound relation becomes somewhat opaque. As C. Chomsky points out, "English orthography represents linguistic knowledge on different levels. In particular, there is a phonological level and a morphological level. The same sound can often be represented by different letters. Which letters are chosen is then decided on a morphological basis: e.g., 'sign' could be spelled sign, syne, cyne, etc. If it relates to 'signature' in meaning, then its spelling must be sign" (1970). Thus, the grapheme-speech relation embedded in the English alphabetic system is characterized as a morphophonemic representation. As a consequence, English orthography is a phonologically deep writing system and the opaqueness of the link between English script and phonology has been seen by many as a barrier to acquisition. Not all alphabetic scripts have such a deep grapheme-phonology relation. For example, Serbo-Croatian, the major language of Yugoslavia, is written in a phonologically shallow

orthography with the simple rule: "Write as you speak and speak as it is written" (Lukatela, Popadić, Ognjenović, & Turvey, 1980, p. 124).

There is an important contrast between logographic and alphabetic scripts with respect to how symbols are packed together to represent the spoken language graphically. For example, in English script, spaces are largely determined on the basis of words: "man," "gentleman," "gentlemanly," "ungentlemanly" and "ungentlemanliness" are each written as a single word even though the last contains five morphemes while the first contains only one. In Chinese script, on the other hand, the spacing is based on morphemes and each morpheme is in fact a syllable: a word like *tricycle* has three morphemes in Chinese (three-wheel-vehicle) and is therefore written with three characters [三 輪 車] and read with three distinctive syllables. Perceptually, the grapheme-sound mapping in Chinese is discrete (i.e., each character is also a syllable) while in English script the relation is continuous and at a more abstract level. This difference may have implications for the beginning readers of these two scripts. For Chinese children, the written array is dissected syllable by syllable and thus has a one-to-one correspondence with the syllables of the spoken language. On the other hand, because of the multilevel representation, a reader of English may have to go through a morphophonemic process in which words are first parsed into morphemes and then symbol-sound relations applied (Venezky, 1970). Furthermore, phonological rules are necessary in order to derive the phonetic form, e.g., to get /sain/ for *sign*. These processes seem very abstract and hence may be quite difficult for a beginning reader.

As we look back at these historical changes, we see that the evolution of writing seems to take a single direction: *At every advance, the number of* symbols in the script decreases and as a direct consequence the abstractness of the relation between script and speech increases. This pattern of development seems to parallel the general trend of cognitive development in children. Results from two independent lines of research are of particular interest. First, anthropological studies (Laboratory of Comparative Human Cognition, 1979) have shown that children's conceptualization of the printed arrays in a text proceeds from pictures, to ideas, to syllables, and finally, to *WORDNESS*. Second, according to E. Gibson (1977), one of the major trends in children's perceptual development is the increasing specificity of correspondence between what is perceived and the information in the stimuli. Similarly, a beginning reader progresses from the whole to the differentiation of the whole, and then to the synthesis of the parts to a more meaningful whole. In a sense, the ontogeny of cognitive behavior seems to recapitulate the evolutionary history of orthographies. This cannot be simply a biological coincidence (Gleitman & Rozin, 1977). Such parallelism implicates the importance of a match between the cognitive ability of the reader and the task demand imposed by the specific orthographic structure of the scripts. One is almost tempted to suggest that orthographic structure in a writing system must somehow mold the cognitive processes of its readers. In fact, it has been claimed that the processes involved in extracting meaning from a printed array depend to some degree on how the information is represented graphically (Besner & Coltheart, 1979; Brooks, 1977; Tzeng & Hung, in press). It is therefore conceivable that different cognitive strategies are required to achieve reading efficiency in various writing systems. One particular concern is whether these different cognitive requirements imposed by various script-

speech relations impose a permanent constraint on our visual information processing strategies, such that readers of different scripts learn to organize the visual world in radically different ways. Evidence for such a new "linguistic relativity" hypothesis can be found in papers discussing the "weak" version of the so-called Whorfian hypothesis (Tzeng & Hung, in press) and in recent ethnographic studies on the behavioral consequences of becoming literate in various types of Vai writing systems (Scribner & Cole, 1978). Cross-language and cross-writing system comparisons are certainly needed to help us answer this and other questions.

Curiously, there has never been a systematic attempt to investigate the effects of orthographic variations on visual information processing. Venezky (1980) characterizes such an absence of studies on orthographic structure as an unfortunate oversight in reading research. He attributes this absence of interest by psychologists in orthography in part to the lack of a linguistic base for describing different orthographic systems and in part to the fact that experimental psychologists in the past were not really interested in the problem of reading. Now the situation has been drastically changed. In 1979 and 1980, three big volumes of theoretical and experimental work on visual language (Kolers, Wrolstad, & Bouma, 1979), spelling (Frith, 1980), and orthography (Kavanagh & Venezky, 1980) were published. In addition, an anthology of experimental work on the perception of print is forthcoming (Tzeng & Singer, in press). It is time to have a critical look at the relation between orthography and visual information processing.

EMPIRICAL DATA

Several points should be clarified. First, although there are many types of alphabetic scripts (English, French, German, Russian, etc.), we will limit our discussion to the English alphabet, mainly because most of the comparative reading studies use English as the representative case. Occasionally, we may discuss other alphabetic scripts when they provide important contrasts to English orthography with respect to certain experimental paradigms. Second, and not unrelated to the first point, most comparative studies have employed the following research strategy: Data and models of processing English orthography are the basic reference points for evaluating data collected with analogous experimental paradigms in non-alphabetic orthographies. Third, the non-alphabetic orthographies here refer to Japanese syllabaries (i.e., kanji and kana) and Chinese logography unless otherwise specified. And finally, in the review itself we assume an information processing approach. That is to say, we first look at studies comparing visual scanning patterns, then at visual lateralization, at some perceptual phenomena such as the Stroop effect, at the issue of speech recoding, at word recognition, and finally, at sentence comprehension. The review is in no way exhaustive and is concerned only with empirical data rather than linguistic speculations.

Visual Scanning

On the surface, the most obvious difference between an English text and a Japanese or Chinese text is that the former is written from left to right and then line by line from top to bottom whereas the latter is usually written from top to bottom and then column by column from right to left. Considering

the fact that most people are right-handed (this is especially true in both China and Japan because of the social-cultural factor that stigmatizes left-handers) and that for right-handed people it is easier to write continuously from left to right, the development of a vertical and right-to-left text arrangement is certainly an unforgivable mistake. The inconvenience can be felt immediately if one attempts to write with a brush and ink. As soon as one moves to the next line, the finished but still wet characters on the right hand side tend to interfere with the current writing unless one consciously lifts the elbow and keeps it in the air all the time. (The ancient Chinese had a special way of training their scholars to be patient and poised.)

Putting aside this inconvenience in writing, is a vertical and right-to-left text easier to read? That is, do we have a natural tendency to scan downward during visual information processing? The anatomical arrangement and physiological structure of our eyes seem to suggest the opposite. Studies in perceptual development have generally found that infants engage in more horizontal than vertical scan (Salapatek, 1968). Moreover, with an equal number of nonsense geometrical figures arranged vertically or horizontally, it has been found that horizontal scanning is quicker than vertical scanning, and this result is observed for both American and Chinese elementary school children. One investigator attributes this difference to the possibility that vertical scanning may result in greater muscular strain as well as quicker fatigue (Tu, 1930). Similar results have also been obtained in Japan with tachistoscopic presentation and with reaction times as the dependent measure (Sakamoto & Makita, 1973). Thus, with respect to reading, there is no evidence suggesting any biological advantage to arranging written text vertically and leftward.[3] The Chinese style has influenced Japanese and Korean text arrangement for centuries, and it is clear that such an arrangement is more a cultural convention than a biological consequence. It is not surprising that a shift toward left-to-right and downward printing has been made in many science texts in order to accommodate Arabic numerals and names of western authors, whose works are usually indexed in the original alphabetic script beside their translations. The readability of such texts seems not to be affected in any systematic way (Chang, 1942; Chen & Carr, 1926; Chou, 1929; Shen, 1927). Our eyes are really very versatile.

It should be pointed out that not all alphabetic scripts are written from left to right. For instance, Hebrew is usually written horizontally from right to left. In fact, in about A.D. 1500 as many scripts were written and read from right to left as from left to right (Corballis & Beale, 1971). Only with the expansion of European culture in later years did left-to-right scripts become predominant. Again, there is no evidence to suggest a biological predisposition for scanning in either direction. Bannatyne (1976) found that eye movement is generally random for 6-year-old or younger French children. However, with older subjects, the left-right eye movements become more or less regular and the regularity increases with the age of the subjects. Apparently, this regularity is a result of reading habit. The following example given by Dreyfuss and Fuller (1972) illustrates this point.

> In South Africa, most of the men who work in the mines are illiterate. The miners, therefore, are given instructions and warning in the form of symbols rather than words. In an effort to enlist the miner's help in keeping mine tracks clear of rock, the

South African Chamber of Mines posted this pictorial message [See Figure 2]. But the campaign failed miserably, more and more rocks blocked the tracks. The reason was soon discovered. Miners were indeed reading the message, but from right to left. They obligingly dumped their rocks on the tracks. (1972, p. 79).

The title of this little example explains the notion very well--"LEFT AND RIGHT ARE IN THE EYES OF THE BEHOLDER."

Although reading direction is merely a learned habit, it seems to have a tremendous effect on the reader's perceptual performance. For example, in one type of speech perception experiment, a subject hears a click while listening to a recorded sentence and is asked to estimate the part of the sentence with which the click was simultaneously presented. With such an experimental paradigm, Fodor and Bever (1965) found incidentally that when the click location task was administered dichotically, the click was judged as coming earlier when it was delivered to the left ear and the speech to the right ear than with the opposite arrangement. Bertelson and Tisseyre (1975) replicated this finding. They conjecture that from the perspective of the subjects, the click is in fact perceived to the left of the sentence, which is presumably transformed into a left-to-right written array. Hence, when the subjects are asked to mark the location of the click on a response sheet, they tend to displace the mark toward the beginning of the sentence, owing to the spatial relation between the click and the sentence. Bertelson and Tisseyre further speculated that the opposite result should be found for Hebrew, which is written from right to left. Indeed, they found that Israeli students, when listening to Hebrew sentences in a similar click experiment, pre-posed the click when the speech was in the left ear and the click in the right more than in the opposite arrangement. Hence, the direction of the effect is inverted when a language that is written from right to left, namely Hebrew, is used in the test. A similar impact of learning to read materials written in different directions (i.e., right-to-left or left-to-right) was also demonstrated on children's visual exploratory patterns. Arrays of pictures of common objects were presented to children who were instructed to name all objects in each array. The exact order of the naming was recorded. While Elkind and Weiss (1967) found a developmental trend of left-to-right directionality in American children, Kugelmass and Lieblich (1979) showed a systematic appearance of a right-to-left directionality in Israeli and Arabic children. These findings are corroborated by Goodnow, Friedman, Bernbaum, and Lehman's (1973) demonstration of the effect of learning to write in English and Hebrew on the direction and sequence in copying geometric shapes.

There also has been some suggestion that the habit of reading direction (i.e., right-to-left vs. left-to-right) affects the pattern of the visual lateralization[4] effect in a visual half-field experiment (Orbach, 1966). We will discuss this issue in more detail in the next section. We mention it here simply as a note on the effect of reading habit on subsequent visual information processing strategies.

We have seen that different arrangements of text in various scripts have a definite effect on reading behavior. In general, horizontal arrangement seems to be more natural from the viewpoint of anatomical arrangement of our eyes and more efficient for writing itself. However, since our eyes are so

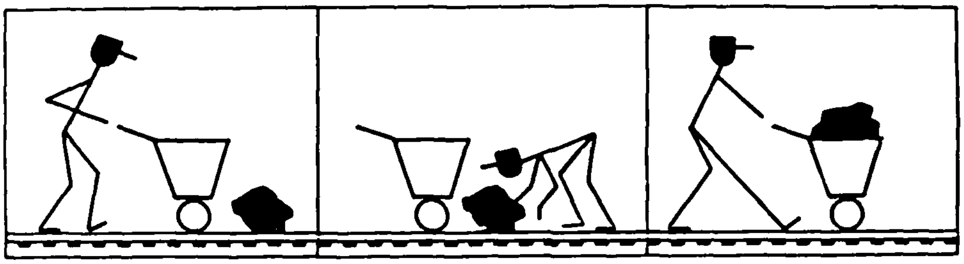


Figure 2. LEFT AND RIGHT ARE IN THE EYES OF THE BEHOLDER (adopted from Dreyfuss & Fuller, 1972).

versatile and flexible, the issue of horizontal versus vertical arrangement may not be too critical. One thing is clear: Once children learn to read the standard style, the pattern of their eye movements becomes stabilized as a result of reading habit.

An important issue has been neglected in all these earlier studies of visual scanning. Very little information is available about the on-line processes during the reading of different orthographic scripts. Since the logographic, syllabic, and alphabetic scripts map onto their respective spoken languages at different levels (i.e., morphemes, syllables, and morphophonemes, respectively), it is important to know whether these orthographic variations affect eye fixation and eye scanning patterns during reading. Such cross-orthography studies of eye movements during reading will no doubt help to resolve one of the key controversies among contemporary investigators of eye movements, namely, the nature and degree of control of individual movements (Levy-Schoen & O'Regan, 1979). For instance, does a Japanese reader tend to skip hiragana symbols based on the knowledge that these cursive scripts usually represent functors in a sentence (as English readers tend to skip THE during reading)? How do Chinese readers compute successive saccadic jumps when word boundaries are not clearly specified in the logographic scripts? Immunity to the effect of such orthographic variations would lend support to the notion of autonomy—that the eyes move to their own rhythm, more or less inflexibly, and with little concern for local variation in the nature of the text. Hence, further research should be directed to basic questions such as the size of perceptual span in each fixation (Rayner, 1978), the number of eye fixations per line given an equivalent amount of information in different orthographies, the length of each fixation as a function of orthographic variations, developmental changes in the eye scanning patterns, and so on.

Neuroanatomical Localization

The human cerebral cortex is divided into left and right hemispheres, and presumably the two hemispheres function cooperatively in normal cognitive activities. However, the idea that these two hemispheres may assume different types of functions was suggested more than 100 years ago (Broca, 1861). Now it is common knowledge that the hemispheres are indeed not equivalent. Sperry, Gazzaniga, and Bogen's (1969) research on split-brain patients provides direct evidence of hemispheric specialization of cognitive function. In these patients, after cutting the corpus callosum (the communication channel between the two hemispheres), the two hemispheres are able to function separately and independently. Sperry et al. (1969) found that written and spoken English are processed in the left hemisphere, while the right hemisphere is superior in performing various visual and spatial tasks. The second line of evidence for this lateralization comes from studies of injuries to the left hemisphere caused by accidents, strokes, tumors, and certain illnesses. These injuries usually impair some language ability, with the kind and degree of the impairment depending on the site and severity of the injury (Lenneberg, 1967; Geschwind, 1970). Evidence for asymmetrically represented functions has also been found in behavioral research with normal subjects. Kimura (1973), for example, found in dichotic listening experiments that subjects were quicker and more accurate in identifying speech sounds transmitted directly (from the right ear via the crossed auditory pathways) to the left hemisphere. Similarly, in visual half-field experiments in which

words were tachistoscopically presented to either the left or the right of a central fixation point, Mishkin and Forgays (1952) found a differential accuracy of recognition, favoring words presented to the right of the fixation point. The last finding has been termed the "visual lateralization effect." It is interesting to note that under certain conditions the visual lateralization effect can also be demonstrated with Chinese-English bilinguals in cross-language testing situations (Hardyck, Tzeng, & Wang, 1977, 1978).

The general pattern that emerges from the results of the above research is the following. In nearly all right-handed individuals and many left-handers as well (Hardyck & Petrinovich, 1977) the left hemisphere is specialized for verbal cognition and memory, including language and most areas of mathematics. The right hemisphere is specialized for nonverbal cognition and memory, including spatial relations and imagery, but also music and other nonverbal sounds. Our concern here is not to review the findings and controversies concerning specialization. Rather, we want to point out that most of these findings came mainly from studies with English or other alphabetic systems. The question is whether orthographic variations make a difference, particularly with respect to data pertinent to reading rather than speech. Evidence has been presented that the nature of the reading impairment depends, in part, on the specific structure of the written language in question (Asayama, 1914). So, our review will focus on the cross-writing-system comparisons of brain-damaged patients and of the visual lateralization effect in normal subjects.

Aphasic Studies in Japan. The major work on the effects of brain lesions on reading Japanese syllabaries has been done by Sasanuma and her associates (Sasanuma, 1972, 1974a, 1974b, 1974c; Sasanuma & Fujimura, 1971). In an earlier review of the literature on reading disorders due to brain lesions in Japan, Beasly (cited in Geschwind, 1971) observed that comprehension of kana scripts is usually more severely affected than that of the kanji script, although the reverse occasionally occurs. Following the implications of this article, Sasanuma has carefully examined the characteristics of the aphasic's speech production and reception and their abilities in reading kana and kanji scripts during and after speech recovery. She reports some evidence for the selective impairment of reading kana and kanji scripts, as suggested by Beasly. Rather than postulating a right and left hemispheric specialization for processing kanji and kana (this dichotomy seems to be implied in Beasly's review), Sasanuma argues for differential disruption of language due to localized lesions in the left hemisphere. The primary difference between reading kana and kanji writings is the necessity of a phonological processor for kana, which is needed to mediate the grapheme-sound-meaning correspondence. It is interesting to note that a similar processor has been postulated for the reading of alphabetic scripts (Rozin et al., 1971). Therefore, Sasanuma's argument has potential for explaining characteristics of language processing beyond Japanese and deserves more careful examination.

Sasanuma has found that most of her patients can be categorized into one of four diagnostic patterns. About half of them had equal impairment for kana and kanji. Another 25% showed the overall symptomatology of Broca's aphasia. On a task that involved writing high-frequency words in kana and kanji, these patients made almost twice as many kana errors as kanji. When asked to write a sentence, they used only kanji characters and the sentence form was similar

to the agrammatical speech of Broca's aphasia. This led Sasanuma to conclude that there was probably a correlation between the impairment of kana processing and an agrammatical tendency. A third group of patients (about 10%) also showed disruption of kana processing, but they differed from the last group in several important respects. In language ability, they were similar to patients with Wernicke's aphasia. A few were diagnosed as having conduction aphasia. These were fluent aphasics as opposed to the nonfluent aphasics with lesions in Broca's area. Their speech was fluently articulated but meaningless. It is also important to note that all patients with selective impairment of kana processing made errors that were phonological in nature. When writing kanji symbols, however, these patients made the same kind of errors as normal subjects--graphemic confusions (Sasanuma & Fujimura, 1971).

The converse was found in the final group of patients who performed better on tasks using kana than kanji. Unfortunately, Sasanuma (1974a) collected in-depth data on only one patient and gave no indication of the prevalence of the disorder. It is apparently a much less common form of aphasia. In writing high-frequency words in kana and kanji, this patient reproduced kana symbols perfectly while missing 80% of the kanji symbols. If he happened to write a kanji character, he used it as if it were a phonetic symbol, without regard for its meaning. Sasanuma classified this patient as belonging to the type of aphasia that has been labeled Gogi aphasia or semantic form aphasia (Imura, 1943) and is similar to the mixed form of transcortical aphasia. This type of patient often can read aloud and dictate in Kana symbols but without any comprehension.

Taken together, these findings would seem to indicate that kana and kanji processing represent distinctively different modes of operation in linguistic behavior. These clinical observations by Sasanuma and her associates are important and provide insights into the mechanisms underlying visual information processing of linguistic materials. Let us summarize these results with some cautious remarks.

1. Most of the aphasic cases reported by Sasanuma and her associates were caused by cerebrovascular accidents. Whenever possible, Sasanuma incorporated reports on neuroanatomical localization into the data. However, it is usually unclear just how precise the localization data are and how secure we can feel about the areas postulated in the aphasic syndromes found in these Japanese patients. Nevertheless, careful examinations of these syndromes and their related reading impairments suggest that these data are consistent with a general pattern of language-specific dyslexic effects reported in other languages (Vaid & Genesee, in press). In general, lesions in the temporal cortex are associated with greater impairment of reading and/or writing of scripts that are phonetically based (de Agostini, 1977; Hinshelwood, 1917; Luria, 1960; Peuser & Leischner, 1974/1980); lesions in the posterior, occipito-parietal cortical areas are associated with greater impairment in reading and/or writing of scripts with a logographic or irregular phonetic basis (Lyman, Kwan, & Chao, 1938; Newcombe, mentioned in Critchley, 1974).

2. There is an odd distribution of the aphasic syndromes, with only one patient with impaired use of kanji and many with impaired kana, which corroborates the disproportional pattern noted by Beasly (see Geschwind, 1971). Thus, the statement of "selective impairment of kana and kanji" may be

misleading. This extremely skewed distribution suggests a totally different interpretation. Rather than hypothesizing differentially localized structures for processing kana and kanji, it might be useful to look at differences in acquisition. One possibility is that kanji characters are difficult to learn and perhaps the long years of practice and special attention spent in learning these characters make them more resistant to loss after brain trauma. This interpretation is interesting but hard to verify empirically. A more attractive interpretation can be offered as follows. The two different pattern-analyzing skills (i.e., recognizing kanji vs. kana scripts) may be viewed as reflecting two different types of acquired knowledge, namely, knowing <u>that</u> versus knowing <u>how</u>. The former represents information that is data-based or declarative, whereas the latter represents information that is based on rules or procedures such as grapheme-sound correspondences (Kolers, 1979). According to Mattingly (1972), operations with these two types of knowledge require two different levels of linguistic awareness. Whereas the realization of knowing <u>that</u> requires only a primary linguistic activity (or Level I ability in terms of Jensen's [1973] classification), the realization of knowing <u>how</u> requires a more abstract secondary linguistic activity (or Jensen's Level II ability). The imbalance between kanji and kana impairments observed in Japanese aphasics may be the result of differential difficulties related to the performance of these two levels of linguistic activities. The dissociation of knowing <u>how</u> from knowing <u>that</u> has recently been demonstrated in amnesic patients (Cohen & Squire, 1980).

3. When discussing the patients that approximate Broca's aphasia, Sasanuma observed a close relation between an agrammatical tendency in speech and an impairment in kana processing. Based upon this observation, she proposed a special phonetic processor and a syntactic processor and further assumed that these two processors were localized close to each other in the left hemisphere. Such a view of dual processors with differential cerebral localizations is suggestive but may be objected to on several grounds. First, that the majority of Sasanuma's aphasic patients were kana-dyslexic. No evidence was provided to show that the kanji-impaired patient was free from the agrammatical tendency. Thus, it is unfair to single out a kana processor. Second, linguistic variations such as kana and kanji scripts do not by themselves justify neurological differentiation unless evidence is provided that rules out other possible interpretations. Third, and more important, there is a more parsimonious explanation that requires no complication of neurological structure. Since the cursive hiragana scripts are used in Japanese writings mainly to represent grammatical morphemes, failure to read hiragana symbols leads directly to the disruption of syntactic structure. Therefore, the close relation between kana impairment and agrammatical tendency should be interpreted as the result of the special function served by kana scripts in Japanese writings.

4. Sasanuma and Fujimura (1971) have reported that Japanese aphasics with apraxia (an impairment of voluntary movement without obvious sensorimotor deficits) of speech perform less well certain tasks requiring visual recognition and writing of kana than do aphasics without apraxia, while the two groups perform comparable tasks with kanji about equally well. The finding that aphasics with apraxia of speech have special difficulty with kana but not kanji is important. Sasanuma and Fujimura (1971) offer the interpretation that apraxic patients have difficulty with the kana script because they cannot

bypass their damaged phonetics and phonology, as they can with kanji. But if the neurological mechanism that is responsible for phonetics and phonology is damaged, then these patients should also show deficiency in analyzing speech. Since these patients did not show any more difficulty in speech perception than the other patients, it is not very plausible to suggest that phonological impairment is responsible for their inability to read kana. Erickson, Mattingly, and Turvey (1977) provide an alternative interpretation. Suppose that it requires more subvocalization to read kana than to read kanji. The apraxic patients would have difficulty in reading kana because of the noise feedback resulting from the imperfect subvocalization. Evidence for more speech recoding activity in reading sound-based scripts such as alphabetic or syllabary scripts has recently been provided by Treiman, Baron, and Luk (1981).

Clinical observations are always very suggestive and should be regarded as a major part of scientific research. However, two apparent shortcomings cannot be avoided in this type of research and were not avoided in Sasanuma's. First of all, the number of cases involved in most clinical studies is usually small; thus, statistical evaluation is difficult. Second, the results are difficult to generalize to normal people. Most clinical observations are collected after the patient recovers from surgical operations. However, little is known about the plasticity of the brain except that reorganization and compensation do seem to occur (Hecaen & Albert, 1978, pp. 394-399). There is also evidence showing that a linguistic task can be accomplished by non-linguistic strategies (Hung, Tzeng, & Warren, in press). Hence, caution should be exercised in making inferences from the recovery patterns of the aphasic patients.

With these comments in mind, let us now turn to the experimental results on visual lateralization effects with normal subjects.

Visual Lateralization Effects. The rationale behind the visual half-field experiment is as follows. When a subject looks at a fixation point in the center of a lighted square within a tachistoscope, each visual half-field projects to the contralateral hemisphere. For example, stimuli presented to the right visual field (RVF) are first processed in the left hemisphere. If language is indeed processed in the left hemisphere, then verbal stimuli presented to the RVF should take less time to respond to than when the same materials are presented to the left visual field (LVF). The delay in reaction time is attributed to the need to transfer information from the right to the left hemisphere. The experimenter can also shorten the exposure duration so that subjects make identification errors. Depending upon the pattern of such an accuracy measure (i.e., RVF or LVF superiority) and upon the materials used, specific functions of the left and right hemispheres can be inferred. With these experimental procedures, most studies have found a RVF advantage for the recognition of English words. This finding is generally referred to as a visual lateralization effect.

Under the influence of Sasanuma's work, investigators have begun to study visual lateralization effects with kanji and kana scripts. When kana symbols are presented first to the LVF and then to the RVF, more errors in a recognition matching task are observed than when they are presented in the reverse order, indicating a left hemisphere superiority for processing kana

script (Hatta, 1976; Hirata & Osaka, 1967). This result is similar to those obtained with alphabetic writings. More recently, Hatta (1977) reported an experiment measuring recognition accuracy of kanji characters and found a LVF superiority for both high and low familiarity kanji characters, suggesting that kanji characters are processed in the right hemisphere. Using a similar experimental procedure, Sasanuma, Itoh, Mori, and Kobayashi (1977) presented kana and kanji words to normal subjects and found a significant RVF superiority for the recognition of kana words but a nonsignificant trend toward LVF superiority for kanji characters. Thus, it seems that for sound-based scripts such as English words and Japanese kana, a RVF-LH superiority effect is to be expected in a tachistoscopic recognition task, whereas a LVF-RH superiority effect is to be expected for the processing of logographic symbols.

The implication underlying this orthography-specific localization hypothesis is that a special phonemic processor is required for the grapheme-sound-meaning mapping in the lexical access of alphabetic and kana words. Although there is indeed evidence for the hemispheric specialization of speech perception (Cutting, 1974; Wood, Goff, & Day, 1971), generalization of such findings to explain the differences between reading logographic symbols and reading alphabetic/syllabic symbols may be misleading. There is now much evidence showing that reading logographic symbols also requires speech recoding under certain circumstances (Erickson et al., 1977; Tzeng, Hung, & Wang, 1977). Thus, the hemispheric difference found in the tachistoscopic recognition of kanji and kana (or alphabetic) symbols reflects, not an orthography-specific localization property but a task-specific property of cerebral hemispheric functioning. To support this claim, Tzeng, Hung, Cotton, and Wang (1979) asked Chinese subjects (all right-handed) to name tachistoscopically presented characters. In the first experiment, Chinese subjects were exposed to brief presentations of single characters in either the RVF or the LVF, and their task was to name the character as quickly as possible. The accuracy data reflected a LVF-RH superiority, replicating previous findings (Hatta, 1977; Sasanuma et al., 1977). Although the results of RH processing are clear cut, its implication for reading is less clear. Modern Chinese tends to be multiple-syllable, and so the perceptual unit in reading may be larger than single characters. Thus, a major task in reading is to generate meaning by putting together several characters to form meaning terms. Recognition of single characters can be accomplished by non-linguistic strategies such as pattern match. Only in combining several morphemes to comprise a meaningful whole does reading require an analytic (linguistic) strategy.

In the second experiment of Tzeng et al. (1979), the stimuli were two characters arranged vertically, and the subjects were asked to name the stimuli (all meaningful terms) as quickly as possible. The procedure of the third experiment was similar to that of the second experiment except that the subjects' task was to decide whether these character strings as a whole were correct semantic terms. (This is a common lexical decision task, and the dependent measure was the reaction time required to make the decision.) A RVF-LH superiority effect was found in both the second and the third experiments. These differential visual lateralization results were difficult to reconcile with the location-specific hypothesis. However, these data are consistent with the view expressed by Patterson and Bradshaw (1975), who assume that the left hemisphere is specialized for sequential-analytic skills, whereas the

right hemisphere performs holistic-gestalt pattern matches. Thus, all these results should be interpreted as reflecting the function-specific properties of the two hemispheres (Patterson & Bradshaw, 1975); they cast doubt on the orthography-specific localization hypothesis proposed by previous investigators. Such a shift of visual lateralization is by no means a unique finding. In fact, Elman (Note 1) reports that even with single kanji characters, a shift from LVF-RH superiority to RVF-LH superiority was observed when the experimental task was changed from simple naming to syntactic categorization (i.e., deciding whether the presented character is a noun, verb, or adjective). A similar shift, though not very pronounced, was also observed in deaf subjects' perception of ASL (American Sign Language) signs (Poizner, Battison, & Lane, 1979). With statically presented signs, a LVF-RH superiority was found; whereas with moving signs, the deaf showed no lateral asymmetry. These latter stimuli included movements of the hands in straight lines; bending, opening, closing, wiggling, converging, linking, divergent, and others. These movements capture much of the significant variation of movement in ASL at the lexical level. Recognition of these movements depends on the ability to put several discrete signs together into a coherent moving sequence. Therefore, the shift from right dominance to a more balanced hemispheric involvement with the change from static to moving signs is consistent with the position that the left hemisphere predominates in the analysis of skilled motor sequencing (Kimura, 1976). It is worthwhile to point out that single ASL signs, like single Chinese characters, sometimes represent morphemes rather than words. In natural signing or in spoken Chinese a meaningful word frequently consists of two or more signs (or characters). The similarity between perceiving ASL signing and reading Chinese characters (despite other differences, cf. Klima & Bellugi, 1979) with respect to the visual lateralization effect strongly suggests that the idea of a left-hemisphere phonetic processor is not viable.

This argument against the orthography-specific localization hypothesis is further reinforced by the observation that procedural differences in a visual half-field experiment may result in either a RVF or LVF superiority effect in the tachistoscopic recognition of Hebrew words (note that Hebrew is an alphabetic script), depending on whether the stimulus words are presented successively in either visual field or simultaneously in both visual fields (Orbach, 1966). Habit of reading direction (right to left for Hebrew) becomes an important factor in this case (Heron, 1957). In fact, all these results are compatible with the substrata-factor theory of reading (Singer, 1962), which asserts that when a task cannot be solved at one level of cognitive operation, a reader may have to fall back on a more analytical mode, perhaps by switching from the right to the left hemisphere. Under this conceptualization, the interaction between orthography and information processing strategy as demonstrated here enables us to identify various subskills at different stages of information processing. The visual lateralization experiment may prove to be a useful technique for untangling this complexity (see Tzeng & Hung, 1980, for a demonstration).

So far, we have reviewed research on effects of orthographic variations on cerebral lateralization using two different approaches, namely, the brain lesion approach and the visual half-field experimental approach. The clinical and experimental studies found differences resulting from reading different scripts, and we have been critical of these findings. However, we do not wish

to deny the existence of these differences. We only argue that these differences can be explained by proposing two types of knowledge (knowing how vs. knowing that) and by the general properties of cerebral organization, without inventing special processors or proposing special locations.

Stroop Interference Experiments

In studies of the Stroop effect (Stroop, 1935), color names are written in an ink of a different color (e.g., GREEN in red ink) and subjects are required to name the color of the ink in which the word is written. In the control condition, subjects name a series of different color patches. It is an established fact that the time it takes to name a series of colors in the test condition is much longer than the time it takes to name a series of color patches in the control condition. Since the Stroop interference effect is very robust and easy to demonstrate, the Stroop task and its variants have been employed by researchers in various fields to investigate different psychological processes, such as the parallel processing of verbal and nonverbal materials (Keele, 1972), the nature of stimulus encoding in short-term memory (Warren, 1972), the properties of bilingual processing (Dyer, 1971; Preston & Lambert, 1969), the automaticity of word recognition in beginning reading (Samuels, 1976), and so on.

A recent study by Biederman and Tsao (1979) with an ingenious application of the Stroop interference paradigm has shed light on the issue of orthographic differences. They observed a greater interference effect for Chinese subjects in a Chinese-version Stroop color-naming task than for American subjects in an English version. They attributed this difference to the possibility that there may be fundamental differences in the perceptual demands of reading Chinese and English. Since, for Chinese characters, the direct accessing of meaning from a pattern's configuration is a function that has been assigned to the right hemisphere, which is also responsible for the perception of color, the increased perceptual load would result in greater interference. For English words, on the other hand, the word processing is mainly a left hemisphere activity; less interference is expected. This study, although intriguing, suffers from several methodological weaknesses. First, there were tremendous subject differences in the reaction times required to name the colors of simple color patches (for some unknown reason, the mean reaction times of the Chinese subjects were relatively slow overall) and differences in verbal ability (i.e., the Chinese subjects happened to be all highly selected graduate students). Second, Chinese color terms are all monosyllabic characters, but this was not true in the case of the English version. Third, all Chinese subjects in the study should be considered semi-bilingual whereas the American subjects were monolinguals. Although Biederman and Tsao did try to rule out the first confounding factor by certain post-hoc statistical analyses and the third confounding factor of bilingualism by citing other bilingual Stroop data, we think that their results should be replicated with a more general subject population.

Shimamura and Hunt (Note 2) and Biederman (personal communication) independently ran the Stroop experiments with Japanese subjects naming the color terms written either in kana or kanji (a within-subject factor). They both found that the kanji version produced more interference than the kana version. Since the same subjects took both the kanji and kana version, the

subject difference was avoided. The result is still consistent with that of Biederman and Tsao (1979). However, a possible flaw may exist in both studies. For fluent readers of Japanese, the color terms they read in everyday life are usually expressed in kanji script and rarely in kana. The greater interference observed for the kanji script may be attributable to this familiarity factor. To counter such an argument, both studies presented further evidence showing that in a simple word-naming experiment (naming words printed in black), color terms written in kana were actually named much faster than color terms written in kanji. Similar findings were reported by Feldman and Turvey (1980). So, although colors are more frequently written in the kanji form and although kanji are more compact graphic representations of words in general, naming time was consistently less for the kana. So far, so good. However, whether one may use naming latency data to resolve the controversy generated by the Stroop task is a question by itself. Since Stroop interference can be obtained in cases where no naming is required (Dyer, 1971), naming speed is hardly an important factor. Thus, although studies of both Biederman and Tsao (1979) and Shimamura and Hunt (Note 2) showed the effect of orthographic variation on the magnitude of Stroop interference, other uncontrolled factors made their data less convincing. Furthermore, with a pictorial variation of the Stroop task, in which subjects were asked to name the pictures as rapidly as possible and ignore the non-congruent words presented simultaneously with the pictures, Smith (Note 3) found no difference in the magnitude of interference between a Chinese version and an English version. This result is opposite to those from studies with colors. One thing that should be noted is that Smith employed multiple-character words, which are linguistically different from the *morpheme*-based single characters used in the color studies. With these ambiguities in mind, let us look at another set of Stroop studies.

In discussing their original finding, Biederman and Tsao (1979) further speculated that there may be some fundamental differences in the obligatory processing of Chinese and English print. They suggested that a reader of alphabetic writing cannot refrain from applying an abstract rule system to the word whereas a reader of Chinese may not be able to refrain from configurational processing of the logograph. Such a conceptualization--that reading different types of scripts may automatically activate different types of perceptual strategies--is intriguing. It leads to a unique prediction concerning bilingual processing in a modified Stroop task. Suppose a Spanish-English bilingual subject is asked to name the color in an English-version Stroop task either in English, the same language as the printed color terms (intra-language condition), or in Spanish, the language different from the printed color terms (inter-language condition). Based on previous empirical findings (Dyer, 1971; Preston & Lambert, 1969), one can predict that the Stroop interference effect should be reduced in the inter-language condition as compared with the intra-language condition. Suppose further that another group of Chinese-English bilinguals are asked to perform a similarly modified Stroop task either in an inter-language or an intra-language condition. Once again one would predict that the Stroop interference should be reduced in the inter-language as compared with the intra-language condition. Of particular interest is the comparison between the Spanish-English and the Chinese-English bilingual subjects with respect to the magnitude of the reduction of the Stroop interference from the intra-language to the inter-language condition. According to Biederman and Tsao's (1979) conjecture that reading alphabetic

and logographic scripts make different perceptual demands, one would predict that the magnitude of reduction should be greater for the Chinese-English bilinguals than for the Spanish-English bilinguals, because English and Spanish are both alphabetic scripts and presumably compete for the same perceptual mechanism (i.e., both would activate obligatorily the same perceptual mechanism for deciphering the alphabetic script). Fang, Tzeng, and Alva (in press) carried out exactly such a modified version of the bilingual Stroop experiment, and the results of their study showed that indeed the magnitude of reduction of the Stroop interference from the intra-language to the inter-language was much greater for the Chinese-English bilinguals than for the Spanish-English bilinguals. This seems to support Biederman and Tsao's contention that reading alphabetic and logographic scripts make different perceptual demands.

Fang et al. (in press) also made an interesting observation. They recalculated from Dyer's (1971) and Preston and Lambert's (1969) bilingual data the magnitude of reduction of the Stroop interference from the intra- to the inter-language condition. All together, there were five types of bilingual subjects: Chinese-English, French-English, German-English, Hungarian-English, and Spanish-English. Fang et al. ranked these bilingual data according to the magnitude of reduction from the intra- to the inter-language condition. The result is as follows: Chinese-English (a reduction of 213 msec), Hungarian-English (112 msec), Spanish-English (68 msec), German-English (36 msec), French-English (33 msec). The ordering of the last three categories is particularly revealing. Why should switching between Spanish and English produce a greater reduction of interference than that between French and English or that between German and English? It is certainly not intuitively obvious why Spanish and English are more orthographically dissimilar than French and English (or German and English). However, if we examine the spellings of color terms across these languages, then the deviation of Spanish becomes immediately clear. For example, red, blue, green, and brown (these colors were used in all these experiments) are translated and spelled as rot, blau, grün, and braun in German; as rouge, bleu, vert, and brun in French; but as rojo, azul, verde, and cafe, respectively, in Spanish. Clearly, with respect to the color terms used in all these studies, Spanish color terms are orthographically more dissimilar to English color terms than both French and German. Correspondingly, the data showed a greater reduction of Stroop interference. The pattern suggests that the magnitude of reduction is a negative function of the orthographic similarity between the two languages involved in the task.

However, since orthographic similarity is highly correlated with phonetic similarity, an alternative explanation for the data is to attribute the effect of switching language to the phonetic factor instead of the orthographic factor. Even though these two explanations are not necessarily mutually exclusive, it is important to determine which factor (orthographic vs. phonetic) contributes more to the reduction of the Stroop interference. To answer this question, Fang et al. ran a similar language-switching experiment with Japanese-English bilinguals. In this case, the pronunciation of the color terms was the same for kanji and kana symbols. If the phonetic factor is responsible for the reduction, then little difference in the magnitude of reduction should be observed between the kanji-English switching condition and the kana-English switching condition. On the other hand, if the orthographic

factor alone can effectively account for the differential reduction, then the magnitude of reduction should be significantly greater for the kanji-English condition than for the kana-English condition. The results of Fang et al. showed that, even with the phonetic factor controlled, the reduction was still greater in the kanji-English switching than in the kana-English switching. Thus, we may conclude that orthographic structure does play an important role, independent of phonological factors, in the lexical access of a bilingual subject.

From the viewpoint of cross-language research, the demonstration of differential perceptual demands in processing different orthographies is an important step toward a general theory of visual information processing. It leads to a host of more intricate questions to be answered. For example, what are these perceptual demands? Do they represent the activation of different knowledge structures (procedural vs. declarative), as speculated in the previous section? Do these differences result in different types of dyslexia? Do they necessitate different instructional strategies for teaching different scripts to beginning readers? To readers learning a second language? Furthermore, does the difference in orthographies (e.g., Chinese-English vs. Spanish-English) also result in different lexical organization? These questions can be answered only by reading research with rigorous experimentation and sophisticated statistical-analytical procedures. Ultimately, we would like to be able to relate the depth of the orthographic structure to the formation of the lexicon in a literate person (either monolingual or bilingual).

<u>Phonetic Recoding in Reading Different Orthographies</u>

Fluent readers can read faster than they can talk, but the opposite is usually true for a child who has just started to learn to read, because the child has to sound out every word in order to get at the meaning. At what point during the process of acquiring reading skills does the transformation of visual code into speech code (a process generally referred to as phonetic recoding) become automatic or even unnecessary (the latter view has been generally referred to as the direct access hypothesis)? The choice between the phonetic recoding hypothesis and the direct access hypothesis has been and still is one of the most controversial subjects of debate in reading research. Experimental data in orthographies other than English are particularly relevant here because of their unique grapheme-meaning mapping rules. For example, the possibility of reading Chinese, in which the logograms do not specify the sound of the word, has been taken as evidence to support the direct access hypothesis.[5] However, a growing number of recent experiments has cast doubt on this general impression of reading Chinese (e.g., Tzeng & Hung, 1980). Let us examine this issue of phonetic recoding versus direct access more carefully with respect to available comparative data.

The idea that readers convert the graphemic representation of printed words into a speech-related code can be traced to the proposal of the subvocalization hypothesis. In its extreme form, this hypothesis asserts that readers must convert the written form into subvocal speech and that, in a sense, reading is no more than listening to oneself. Although there is evidence supporting this hypothesis (Hardyck & Petrinovich, 1970), a moment's reflection suggests it can easily be refuted on both logical and empirical

grounds. For one thing, it asserts that a fluent reader can never read faster than he can talk. This we already know is not true. Second, Rohrman and Gough (1967) and Sabol and DeRosa (1976) have shown that subjects can gain access to a word in the mental lexicon in less than 200 msec, whereas naming a three-letter word requires approximately 525 msec (Cosky, 1975). Thus, it is absurd to assert that readers have to wait to receive subvocal information before they gain access to the lexical memory of words.

The phonetic recoding hypothesis differs from the subvocalization hypothesis in that the grapheme-speech conversion is at a more abstract level, thus avoiding the tedious motor process of vocalization. There is a great deal of evidence that phonetic information is often used during the decoding of written English. In the early 60's, researchers on memory accumulated much evidence suggesting that phonetic recoding occurs in processing verbal materials even if they are presented visually (Conrad, 1964). These experiments generally found that confusion in short-term memory is more often due to phonetic similarity between the to-be-remembered and the interpolated items than to visual or semantic similarity. Analysis of the kinds of errors the subjects make suggests that a grapheme-speech code conversion occurs and that this speech code is phonetic in nature (Baddeley & Hitch, 1974).

Another source of evidence for the phonetic recoding hypothesis is work by Corcoran (1967) and others who have demonstrated that spelling errors resulting in a letter string that is pronounced like a word go undetected more often than errors leading to letter strings that do not sound like words. Similar results were obtained by MacKay (1972) with a different experimental paradigm. These investigators have taken these data to suggest that the reader has translated the printed words into a phonetic representation that corresponds to an entry in his mental lexicon such that the spelling errors go undetected.

Considerable evidence has been accumulated that shows a syllable effect in reading-related tasks: disyllabic or multisyllabic words are named more slowly than monosyllabic words; same/different judgments are slower for multiple-syllable than single syllable items, and letter detection is more accurate in monosyllabic than disyllabic words (see Massaro, 1975, for a general review). Since the syllable effect is obtained for words equated for visual length, the effect can be taken to indicate translation into a phonetic form during the visual recognition process. However, one should take extreme caution in interpreting results of a naming task. At least two processes should be distinguished: (1) visual recognition and (2) articulating the response. A syllable effect can be localized in either process, but our theoretical interest is in only the first, since our concern is really with how speech is used to gain access to meaning during the initial contact with print. An experiment that demonstrated the syllable effect without the contamination of the naming process (Pynte, 1974) is particularly revealing in this connection. Pynte found that French people gazed longer at two-digit numbers whose names contained more syllables (e.g., 82 is pronounced as quatre-vingt deux, with four syllables) than at those whose names contained fewer syllables (e.g., 28 is pronounced as vingt huit, with only two syllables). The syllable effect observed in reading numbers is important because Arabic numerals are logographic symbols and it has been assumed that reading logographic scripts does not engage any phonetic recoding. Apparently, this assumption is not valid.

Experiments using lexical decision tasks provide a fourth source of evidence in favor of the phonetic recoding hypothesis. Rubenstein, Lewis, and Rubenstein (1971) presented letter strings to their subjects and simply asked whether or not each letter string was an English word. They found that subjects took considerably longer to reject pseudowords that are homophonous with (sound like) real English words (e.g., brane) and that nonpronounceable items (e.g., sagm) were rejected most rapidly. In another experiment, these investigators also found slower positive responses for words that are homophonic in nature, such as yoke-yolk and sale-sail than for control words such as moth. Meyer and his associates (Meyer & Ruddy, Note 4; Meyer, Schvaneveldt, & Ruddy, 1974) have replicated and extended these findings to experimental situations involving lexical judgments of pairs of letter strings.

In summary, a number of experiments using a variety of techniques have produced evidence that the phonological structure of a word affects its visual processing. This evidence is consistent with a phonetic recoding hypothesis. However, the seemingly clear picture becomes muddied when we begin to examine other sets of experimental results, which support the direct access hypothesis, that readers are able to go directly from the graphemic representation of the printed word to the lexical representation in their mental dictionary.

First, Baron (1973) demonstrated that subjects had no more difficulty in deciding that a phrase was nonsense when it sounded sensible than when it did not. For example, they could classify the phrase, TIE THE NOT, as nonsense, as quickly as the phrase, I AM KILL. According to the phonetic recoding hypothesis, one would have expected the phonemic correctness of the first phrase to slow down rejection time if phonetic translation had indeed occurred. But this expectation was clearly not confirmed. Second, Bower (1970) asked speakers of Greek to read passages containing misspellings that were pronounced exactly the same as the correct spellings. This was accomplished by interchanging vowels that were pronounced identically but spelled differently. The Greek readers were considerably slowed down by this visual distortion, suggesting that their normal reading must be via some route disrupted by the visual change. Obviously, the grapheme to phoneme route was still available and undistorted (though it was less familiar), indicating that it was not the only route used during rapid reading. Third, Davelaar, Coltheart, Besner, and Jonasson (1978) have shown a dependence of the homophone effect on the exact items used in the lexical decision judgments. In their experiment, Davelaar et al. included one comparison (MOTH vs. YOKE) under Rubenstein et al. conditions, with nonwords like SLINT. The result showed a reliable slower response time for YOKE than that for MOTH (628 vs. 606 msec). When the experiment was changed slightly by including nonwords (like BRANE) that were homophonic with real words, the previous difference in response time between YOKE and MOTH (600 vs. 596 msec, respectively) went away. The conclusion seems clear: an optional, not compulsory, speech-based process is involved in lexical access and the subjects can bypass it when the task demands make it a poor strategy.

A final but perhaps the strongest set of evidence against the phonetic recoding hypothesis comes from an experiment conducted by Kleiman (1975). Kleiman presented subjects with a pair of words and asked them to make one of three types of judgments: (a) graphemic similarity, (b) phonemic similarity, and (c) semantic similarity (synonymity). On some trials the subjects were

also required to "shadow" a series of digits heard through an earphone while performing the judgment task. On other trials they performed only the judgment task. Kleiman found that prevention of phonemic translation had little effect on graphemic and semantic judgments as compared with performance on phonemic judgment. Since semantic judgment required access to meaning, this result suggests that meaning access does not depend on grapheme-phoneme conversion.

We have seen evidence for and against the phonetic recoding hypothesis with respect to the reading of alphabetic materials. What about parallel lines of research in reading logographic materials? In fact, supporters of the direct access hypothesis have always used the example of reading Chinese logographs to reinforce their argument. The argument goes like this: Since Chinese logographs do not contain information about pronunciation, people must be able to read without speech recoding. This statement is not exactly correct. First of all, the majority of Chinese logographs are phonograms that at times do give clues to the pronunciation of the character (the efficiency coefficient for correctly predicting pronunciation of a phonogram from its constituent sound component is estimated to be .36, see Tzeng & Hung, 1980). Second, reading should not be equated with lexical access of a single word; rather, it should be regarded as a more general linguistic activity that involves all sorts of subcomponent activities such as iconic scanning and storage, lexical retrieval, short-term memory, syntactic parsing at both the macro- and micro-levels (Kintsch & Van Dijk, 1978), and semantic integration (Bransford & Franks, 1971). This kind of conceptualization immediately questions the validity of the view that reading logographic script such as Chinese involves no grapheme-phoneme translation. Such translation may not be necessary at the entry of the lexicon, but it may very well occur during the short-term memory stage or the syntactic parsing stage.

Tzeng et al. (1977) carried out two experiments to investigate whether phonemic similarity affects the visual information processing of Chinese characters. The first experiment employed a retroactive interference paradigm introduced by Wickelgren (1965). Chinese subjects were asked to memorize a list of four unrelated characters presented visually followed by the shadowing of a series of aurally presented characters that were phonemically similar or dissimilar to the target characters. The results showed a tremendous amount of intralist and interlist interference due to phonemic similarity. This is consistent with the experimental results in English (Conrad, 1964; Kintsch & Buschke, 1969; Wickelgren, 1965). Furthermore, vowel similarity produced more interference than did consonant similarity. This finding is consistent with previous experiments by Crowder (1971) with alphabetic materials and a very different experimental procedure. In their second experiment, Tzeng et al. extended the finding of such a phonemic similarity effect to a sentence judgment task. The experimental task required subjects to judge whether a singly presented sentence was a normal sentence or an anomalous sentence. Normal sentences were both grammatical and meaningful whereas anomalous sentences were both ungrammatical and relatively meaningless. The major independent variable was the degree of phonemic similarity among the characters that made up the sentences; the dependent measure was the reaction time required for making a correct judgment. The results clearly showed that performance in such a sentence judgment task was impaired by the introduction of phonemic similarity into the test material. Erickson et al. (1977) also

demonstrated the effect of phonemic similarity with Japanese subjects memorizing a list of kanji characters.

In another experiment, Tzeng and Hung (1980) asked Chinese subjects to read a section of prose containing about 1500 characters and concurrently circle all characters containing certain graphemic components such as 台 or 由. These two graphemic components sometimes are used to construct phonograms but sometimes they have nothing to do with the pronunciation of the entire character. For example, the pronunciation of 胎, /tāi/ is based on the sound of 台, /tái/ while that of 怡, /í/ is not, even though both characters contain the same graphemic component 台 on the right-hand side. It was found that subjects detected more characters in which the designated graphemic component carried a phonetic clue. This result is similar to Corcoran and Weening's (1968) finding that when English-reading subjects are asked to perform a similar task, they detect the embedded letter e more often when it is sounded than when it is silent. One may argue that since the findings reported by Tzeng and his associates were obtained with Chinese students who are to some extent bilinguals, the results may be attributed to their having been exposed to alphabetic materials. This argument was weakened by a recent study with Chinese children who had just started to learn Chinese characters. Chu-Chang and Loritz (1977) found that in a Chinese character recognition task, where a tachistoscopically presented character list was followed by a list consisting of corresponding phonological, visual, and semantic distracting characters, the children responded predominantly to phonological distractors.

To explore further the contrast between processing logographic and alphabetic scripts with respect to the issue of phonetic recoding, Tzeng and Hung (1980) ran an experiment similar to that of Kleiman (1975). They asked Chinese subjects to make one of four types of judgments about two simultaneously presented characters that were flashed very briefly in the tachistoscope: (a) graphemic similarity (share an identical radical), (b) phonemic similarity (rhyme with each other), (c) semantic similarity (synonymity), and (d) sentence anomaly (grammaticality of a sentence). Again, on some trials subjects were concurrently engaged in a digit shadowing task while performing the decision task and on other trials they were not. Tzeng and Hung found that the phonemic decision was seriously affected by the shadowing task, whereas both the graphemic and semantic decisions seemed to suffer only from general disruption caused by the shadowing task. The authors concluded, like Kleiman with his data on English, that lexical retrieval of single characters does not require any grapheme-phoneme translation. Of particular interest was the result of the sentence-judgment condition. It was found that sentence judgment was also affected greatly by the shadowing task, suggesting a performance impairment caused by the prevention of the grapheme-phoneme conversion.

One implication to be drawn from all these findings is that phonetic mediation is just one of the strategies for obtaining access to meaning, rather than an obligatory stage. The use of phonetic recoding may depend on such factors as the difficulty of the materials and the reader's purpose (e.g., whether he wishes to commit the material to memory). Hence, Tzeng et al. (1977) concluded: "There are at least two major ways in which phonetic recoding is claimed as an important process in reading. First, in blending

the individual letters of words, the phonetic recoding of the individual letter sound can plausibly be argued as an important intervening stage, at least for children learning to read. A second way in which phonetic recoding may be involved in reading is concerned with the question of whether fluent adult readers need to phonetically recode printed material or are assisted by doing so. In this latter view the phonetic recoding is viewed as a general strategy of human information processing, and thus the orthographic difference in the printed materials becomes less important" (p. 629). The view that the role of speech in lexical access changes with increasing experience in reading was confirmed in a developmental study by Barron and Baron (1977). They reasoned that at the beginning stage of reading, children may need to sound out words in order to match them with the only lexical system they have at the time, a lexical system organized by speech; however, as fluency develops, direct connections emerge between the printed words and their meaning, resulting in a visually-organized lexicon. Barron and Baron's experimental results were consistent with such a dual-lexicon hypothesis. This tendency of shifting from a speech-based lexicon to a visually based lexicon seems to be a universal phenomenon of fluent reading behavior. Based upon clinical observations of Japanese aphasic patients, Asayama (1914) suggested that the "sensory-acoustic" center of the cerebral cortex plays a major role in the initial learning of kanji because it is not acquired ostensively but rather by way of the oral Japanese translation. With practice and experience, the significance of this center diminishes until, finally, associations between the "optic center" and the "concept center" can take place directly without involvement of the sensory-acoustic center. Thus, a general principle seems to hold for fluent readers regardless of whether the scripts contain sound-based symbols or morpheme-based logographs--a speech code may not be necessary for lexical access, but it is certainly useful for short-term memory. This conclusion is similar to the one reached by Liberman, Liberman, Mattingly, and Shankweiler (1980), that the requirement of a phonetically based working memory for linguistic comprehension should be a universal phenomenon.

Before we leave the debate on the phonetic recoding hypothesis versus the direct access hypothesis, let us remember Campbell and Stanley's (1963) admonition about opposing theories. "When one finds...that competent observers advocate strongly divergent points of view, it seems likely on a priori grounds that both have observed something valid about the natural situation. The stronger the controversy, the more likely this is" (p. 3). Campbell and Stanley's observation certainly applies to the phonetic recoding versus direct access issue in reading.

Given the possibility of two different paths leading from the print to the two lexicons (speech-based or visually based), the existence of some speech recoding activities is no longer in doubt. The question now facing us is <u>when</u> they are used. What factors encourage their use and what factors discourage it? Undoubtedly, study of the different forms of script-speech relation--Chinese logographs, Japanese syllabaries, vowel-free Hebrew, and so on--should reveal further constraints upon possible patterns of speech recoding during reading. For example, English and Chinese writings differ along an important dimension: the extent to which one can predict sound from the printed array. It is quite possible that differences in orthographies along this dimension affect the use of speech recoding in silent reading. If the written forms on the page stand in a regular relation to the sounds of

language, readers may use the grapheme-sound rules to help them derive the meanings of words. Such a path would be largely unavailable to the readers of Chinese but would be highly available to English readers. Therefore, one may expect readers of English to engage in speech recoding more than would Chinese readers. A recent experiment comparing the degrees of speech recoding between Chinese and English readers confirmed this expectation (Treiman et al., 1981).

One can push the argument even further and make the claim that in an alphabetic script where the prediction of sound from letters alone is always valid (i.e., a perfect spelling-to-sound regularity), readers may automatically activate the phonological route to the lexicon. Experiments with a phonologically shallow orthograhy such as Serbo-Croatian (the major language of Yugoslavia, which can be written in either Roman or Cyrillic) have consistently demonstrated that lexical decision proceeds with reference to the phonology (Lukatela et al., 1980). Most important, these investigators found that even when matters were arranged so as to make the use of a phonological code punitive in accessing the lexicon, readers of Serbo-Croatian were unable to suppress the phonological code. This result is directly opposite to that obtained with English. Davelaar et al. (1978) found that under similar arrangements, readers of English abandoned the phonological route and opted for direct visual access to the lexicon. Thus, in a less shallow orthography such as English, reading may proceed simultaneously at several levels of linguistic analysis. The concept of depth with respect to the orthographic structure seems to be a useful construct in evaluating the issue of speech recoding.

From the above discussions, there is an interesting speculation to be made. In between Serbo-Croatian orthographies, which have excellent letter-sound correspondences, and Chinese logography, which has only very fuzzy sound clues, we have other orthographies such as English, which are phonologically deep and thus are graphemically and phonemically opaque. According to Baron and Strawson's (1976) classification of Phoenician (those who attend to the phonetic aspects) and Chinese (those who attend to the visual aspect) readers, one should expect that fluent readers of Serbo-Croatian are disproportionately Phoenician and fluent readers of logography are disproportionately Chinese. For fluent readers of English the proportions of Phoenician and Chinese should be roughly equal with a tendency of being skewing toward becoming more and more Phoenician (Lukatela et al., 1980). It seems that the development of coding options and the development of meta-cognitive ability in order to optimize certain coding strategies relative to appropriate linguistic contexts are essential for becoming skilled readers of a phonologically deeper orthography such as English. Here is an area in which comparative reading studies across different orthographies can yield important information.

Word Recognition

The processes by which words are recognized in isolation have occupied the attention of many experimental psychologists over the last hundred years. Research in this area has made significant contributions to our understanding of pattern recognition, memory structure, the relation between speech and reading, and cognitive functioning in general. However, cross-language studies, especially cross-writing-system comparisons of word recognition processes, are very much needed. The reason is simple and straightforward.

Different orthographic structures exhibit different script-speech relationships, and perceptual pathways leading from print to meaning seem to be constrained by these differences, as shown by different degrees of speech recoding activity and different patterns of Stroop interference. It should also be pointed out that current models of word recognition such as Morton's (1969) logogen model and the spreading activation model of Collins and Loftus (1975) make the assumption that orthographic information is contained in semantic memory. This assumption was verified in a recent study by Seidenberg and Tanenhaus (1979) by the demonstration that the orthographic code is readily available even in an auditory word recognition task. They showed that in a listening experiment, subjects were markedly slower in deciding that "rye" and "tie" rhyme than that "pie" and "tie" do. Thus, by examining factors that affect word recognition in different writing systems we should be in a better position to specify the nature of logogens in our semantic memory.

In general, it seems that similar factors affect recognition of logographic characters and of alphabetic words. Solomon and Postman (1952) demonstrated that in English the recognition threshold for high-frequency words is lower than for low-frequency words. Other variables that also influence word recognition include meaningfulness (Broadbent, 1967), imagery, and concreteness (Paivio, 1971), with higher value in these dimensions being associated with lower thresholds. In Chinese, these same variables also show similar effects on character recognition. Yeh and Liu (1972) demonstrated the effects of frequency and meaningfulness on the recognition threshold. The effectiveness of imagery and concreteness were substantiated by the experimental work of Huang and Liu (1978). One interesting observation should be noted here. In English, word length has been found to be a negative function of frequency of usage and this has been referred to as one type of Zipf's law. The same observation seems to hold in the case of Chinese characters. Thus, whereas the average word length in English is about five to six letters, the average number of strokes in common Chinese characters is about six (Wang, 1973). In both cases, the graphemic development seems to favor the direction of perceptual ease and production economy. In another interesting study, Nelson and Ladar (1976) selected randomly a list of characters from norms of scaled meaningfulness in Taiwan (Liu & Chuang, 1970) and asked Canadian college students who had no experience with Chinese to rate these characters for their visual meaningfulness. The result showed that the amount of perceptual information in these characters as conveyed to those English-speaking observers correlated significantly with the index of associative meaningfulness for Chinese-speaking individuals. Similar studies were also carried out by Koriat and Levy (1979) who showed that Israeli students noncognate of Chinese were able to correctly guess the meanings of Chinese logographs with better than chance success.

Psychological studies such as these can yield insights as to how characters evolve through the years. In order for such a correlation to hold, one has to assume that, on the one hand, high frequency of usage has forced simplification of the characters and, on the other hand, the graphemic simplification and formalization process is constrained by universal perceptual-motor factors. The first assumption is easy to defend, but the second assumption deserves critical analysis. In a recent study, Tzeng, Malley, Hung, and Dreher (Note 5) demonstrated that even in simple drawings of common objects, such as a coffee cup, people tend to exhibit the history of their

interaction with the object. For example, most people draw a coffee cup with a handle on the right-hand side because that is the way they usually hold the cup. The argument advanced by Tzeng et al. is that graphemic information is subject to certain perceptual-motor constraints. If such is the case, then visual recognition of Chinese characters may be aided by such constraints, just as the Canadian and Israeli students are able to gather some meaningful information from the graphemic information alone. All these results suggest that choice of orthographic code to designate concepts is not arbitrary but is rather governed by lawful, cross-culturally consistent, figural-semantic associations (Koriat & Levy, 1979).

Another important research topic in current word recognition studies concerns the issue of the so-called word superiority effect (WSE). Almost a hundred years ago, Cattell (1886) discovered that with very brief exposures a letter can be reported more accurately when it is embedded in a word than when it is presented alone. Since then, this WSE has been repeated and confirmed. Reicher (1969) performed an experiment that rules out a simple guessing theory. Immediately after exposure of a stimulus word, Reicher tested one critical letter position with a forced choice between two alternative letters (e.g., a choice between "D" and "K" after the word "WORD"). The key to the experiment is that both critical letter alternatives always made a word in the context of the other stimulus letters (e.g., "WORD" and "WORK"); in fact, each letter alternative was equally likely to appear in the presented context. To measure the WSE, the same critical letter was presented in an unrelated letter string (e.g., "RWOD") again, followed by a forced choice between "D" and "K" as alternative last letters. Reicher (1969) found that performance for a letter in a word was substantially higher than for a letter in an unrelated letter string, and indeed higher than for a single letter presented alone.

A number of investigators soon pointed out that a modified version of the sophisticated guessing theory could be formulated to account for the WSE obtained with Reicher's paradigm (for a review, see Johnston, in press). Experiment after experiment was conducted to set up the parametric boundary of this effect. In fact, the WSE has become one of the most important experimental paradigms in evaluating theories of word recognition. It is not our intent to review all the theories and models constructed to explain this effect; but we would like to highlight two contrasting views of the WSE and review a study of this effect with kana symbols that helps to clarify these two contrasting views.

One important observation on the WSE is that the superiority effect is not restricted to meaningful words. It can readily be demonstrated with pseudowords that follow the orthographic regularities of English spelling. Since orthographic regularity is correlated highly with pronounceability, the observed superiority effect has usually been attributed either to the orthographical regularity of the letter groups (e.g., Massaro, 1975) or to their syllabic nature (Spoehr & Smith, 1975). The latter view is called the vocalic center group (VCG) hypothesis, according to which a syllable-like structure is the perceptual unit for word recognition. The reason for the superiority in the perception of words and pseudowords is that the perceived letter strings are readily parsed into VCGs.

The VCG hypothesis has recently been challenged by a study in Japan (Miura, 1978) that demonstrated a WSE with kana script using Reicher's experimental paradigm. Since each kana symbol has an invariant one-syllable pronunciation, the superiority effect obtained cannot be attributed to the advantage of parsing into a VCG. Actually, a VCG model would predict that word and nonword recognition accuracy should be the same and should be lower than for the single kana symbol. The results were just the opposite of these predictions. Miura therefore suggested that a model based upon orthographic regularities may be a better candidate for the interpretation of the WSE. Unfortunately, no corresponding experiment on the WSE has been run with Chinese logographs. It would be extremely interesting to make such a cross-orthography comparison. We mentioned that the WSE could be obtained with pseudowords. One could make counterfeit Chinese characters and see if the WSE still occurred for Chinese readers. Maybe the locus of the WSE lies neither in the speech pathway nor in the visual pathway to the lexicon but in the memorability of a more abstract and integrated code, as recently suggested by Johnston (in press).

In recent experimental work with English materials there is another interesting finding: For English-speaking subjects, written words are named markedly faster than pictures of common objects but are classified by meaning (semantic categorization task) more slowly than pictures (Potter & Faulconer, 1975). The difference cannot be readily explained by uncertainty as to the name of a pictured object or by features that allow pictures to be classified without full recognition. The general pattern of these results suggests that a picture of an object and its written English name ultimately activate in memory one and the same concept or meaning, accounting for the near equality for pictures and words in classification time. A word, however, appears to activate an articulatory mechanism before activating its concept, so that written words can be named rapidly. For a pictured object, access to the articulatory mechanism is apparently indirect; the object's concept must be activated first and then the associated name retrieved, so that naming is slow. Thus, the status of words and the status of pictures are experimentally differentiated.

One challenging question has always been raised with respect to the recognition of Chinese logograms: Is the recognition process more similar to picture perception or to word recognition? This distinction is similar to Huttenlocher's (1975) distinction between "reference-field schema" and "symbol schema" and has been shown to be linguistically meaningful in differentiating sign language from spoken language. Many linguists and reading specialists (Gibson & Levin, 1975) have speculated that Chinese logograms are similar to pictures and different from English words in three respects: They are graphically unified, they may represent features of their reference directly (e.g., the trunk and branches of a tree make up the character for wood, 木), and they do not represent the component sounds of their spoken names. On the other hand, logograms are also like written English words and different from pictures in that they, as symbol schemata, relate to the reference field only indirectly through encoding and decoding processes. Thus, a comparison with picture perception may indicate whether the pictorial properties of Chinese characters or their status as words determines how they are processed. If processing Chinese logograms is more like picture perception, then one would expect that Potter and Faulconer's (1975) experimental procedures would yield

smaller differences between logograms and pictures in naming and classification tasks for Chinese readers, compared with the pattern obtained with English readers.

Two experiments were carried out by So, Potter, and Friedman (Note 6) on the time it takes Chinese subjects to name logograms and to classify them according to meaning. For the purpose of cross-language comparison, they also reran the experiments with English subjects naming or classifying pictures and words. The results showed that in English as well as in Chinese, written words are named faster than pictures. The magnitude of the difference is almost identical in both languages. So, contrary to the speculation that written Chinese is harder to pronounce and easier to understand than written English, both languages are very similar in the processing of information. This finding for Chinese and English suggests that in any language there is a direct link between a written word and its spoken name, even when the writing system does not represent the component sounds of words.

The question of whether Chinese logographs are processed like pictures was also tested with a picture-word interference paradigm (Smith, Note 3). In a pictorial variation of the Stroop task, subjects are presented with a series of line drawings, each containing a noncongruent word. For example, a drawing of a chair may contain the word "hat." Subjects are asked to name the pictures as rapidly as possible, ignoring the words. Typically, the presence of an incongruent word results in considerably slower naming time compared with a control condition in which pictures are presented without words (Rosinski, Gollinkoff, & Kukish, 1975). Smith (Note 3) reasoned that if Chinese words are processed like pictures, then more interference should be observed with Chinese readers than with French readers in a similar picture-word interference task. Her results were negative, suggesting that words written with Chinese characters are no more processed like picture than words written with alphabetic scripts.

According to the logogen model (Morton, 1969) and the semantic-network model (Collins & Loftus, 1975) of word recognition, the linguistic unit with which the logogen or concept is concerned is, roughly, a word. We have mentioned that a single Chinese character should not always be equated with a word. For example, the English word library is written as a three-character compound, 圖書館, in Chinese. Thus, the word is a more abstract code, compared with a single character. It is no wonder that at the level of the word, the logogen should be independent of the orthographic factor. Factors such as frequency of usage, imagery, meaningfulness, and concreteness are concerned with the logogen itself. So these factors should have similar effects on words written in different orthographies. Only factors that specifically concern the connection between print and the logogen should show differential effects on word recognition in different orthographies. Besner and Coltheart (1979) asked their subjects to choose the larger number from a pair of digit numbers printed in different sizes, and found that subjects' choice reaction times were subject to the interference of size-incongruency (e.g., when the symbol for 6 was much larger than that for 9) only when the numbers were presented in Arabic numerals (i.e., logographic symbols) but not when they were presented as spelled-out English words (i.e., SIX vs. NINE). Apparently, different mechanisms are involved in making the connection between print and the logogen in these two cases. So, with respect to results of

different types of experiments on word recognition, the conclusion to be drawn is that at the level of the word, the orthographic variation does not seem to matter much. At the level of words, script and speech converge on an amodal linguistic entity.

Sentence Comprehension

We have reviewed so far the effects of orthographic variations on visual information processing from the most superficial level of eye scanning to the deeper level of word processing. We have found that processing differences for different writing systems seem to occur at the lower level, with little difference beyond the level of the word. Our attention will now shift to sentence processing. Ordinarily, real-life reading involves comprehension of individual sentences as well as integration of semantic contents across paragraphs within a text. We would not expect to find any processing difference due to different orthographies at such higher level processing. Although there have not been many studies on this issue, our general impression based on currently available data is that similarity seems to be the rule across different orthographies.

Just and Carpenter (1975) employed the picture-sentence verification paradigm to examine sentence comprehension in Chinese, Norwegian, and English. This experimental paradigm was first established by Clark and Chase (1972), who asked their subjects to decide whether a sentence was true or false according to an accompanying picture. For example, if a sentence is IT'S TRUE THAT THE DOTS ARE RED and the picture is of red dots, subjects' response should be "Yes," and this sentence is classified as a true affirmative (TA) sentence. If the picture shown is of black dots, then subjects would respond "no," and the sentence is a false affirmative (FA) sentence. There are also negative sentences. For instance, if the sentence is IT'S TRUE THAT THE DOTS ARE NOT RED and the picture is of black dots, then this is a true negative sentence (TN) and subjects' response should be "yes." Again, if the picture is red and the subjects' response should be "no," the sentence is a false negative (FN) sentence. Based upon an analysis of the verification process in each case, Clark and Chase (1972) were able to predict that the verification times for the four types of sentences should be ranked as TA<FA<FN<TN.

Carpenter and Just (1975) further elaborated and modified the Clark and Chase (1972) model and developed the so-called constituent model of sentence verification. This model assumes that all internal representations, whether of pictures or sentences, are propositional. The verification processes start at the most inward constituent propositions. For example, the TA sentence can be represented as {AFF(RED,DOTS)}. Since the picture is also represented as (RED,DOT), the time it takes to compare the sentence with the picture should be the quickest because of the direct match (the time required to do the comparison is called k units of time). Whenever corresponding constituents from the sentence and picture representations mismatch, the comparison process is reinitiated, so the total number of comparison operations, and consequently the total latency, increase with the number of mismatches. Accordingly, the time it takes for the verification of FA sentences will be k+1 since an additional mismatch has been found. The FN sentence is represented as {NEG(RED,DOTS)}, which results in two additional mismatches; thus it should take k+2 units time to verify. The propositional representation for a TN

sentence is {NEG(RED,DOTS)} but the picture is represented as (BLACK,DOTS). Therefore, three additional steps are required in this case in order to be able to verify the sentence; consequently it takes $k+3$ units of time. (For detailed analysis of these verification times, see Carpenter & Just, 1975.) This model predicts beautifully the sentence verification times for these four types of sentences.

With this experimental paradigm, Carpenter and Just (1975) ran two cross-language experiments and fitted their model to the data. In their first experiment, they used Chinese subjects and all sentences were written in Chinese. They found a remarkable similarity between sentence verification processes in Chinese and English even though word boundaries are clearly defined by spacing in printed English sentences but not in printed Chinese sentences. The time per constituent comparison (i.e., k), 210 msec for Chinese sentences, is very close to the 200 msec for English. Thus, processing rates and modes of processing are similar even though these two languages come from very different language families and even though these two writing systems represent their respective spoken languages at very different levels.

In Carpenter and Just's second experiment, the same procedures were used to test Norwegian subjects with sentences written in Norwegian. One complication was added: a quantifier variable was included. For example, the sentence was IT'S TRUE THAT MANY (or A FEW) OF THE DOTS ARE RED. They found that mean latencies increased with the number of constituent comparisons for both kinds of quantifiers. The processing time per operation was slightly longer in Norwegian (322 msec in the first block of testing and 278 msec in the second and third blocks), compared with those of English and Chinese (200 msec and 210 msec, respectively). However, there were fewer practice trials and sentences and pictures were more complex in this experiment.

Overall, there seems to be considerable universality in the underlying mental operations across three languages. It is of particular interest that the time for each additional retrieval and comparison in this type of task is very close to the duration of the scanning and comparing operation (240 msec) found by Sternberg (1969) in a context recall experiment. This suggests that a common fundamental operation underlies different tasks, across different languages. It is worthwhile to mention that in a recent study with a similar sentence-picture verification paradigm, Hung, Tzeng, and Warren (in press) found that deaf subjects engaged identical schemes to process signed sentences. Such commonalities point toward an explanation of language universals through the discovery of processing universals.

The experiments just mentioned have monolingual subjects processing sentences written in their own languages. What would happen if bilinguals were to read materials written in mixed languages? Do they use a dual linguistic system or a single cognitive system but with specific linguistic information stored at some points? Tsao (1973) used Chinese-English bilingual subjects to study this issue. He employed Bransford and Franks' (1971) experimental paradigm to investigate the abstraction and integration of ideas across sentences when sentences were presented all in Chinese or all in English (the single-language condition), or half in Chinese and half in English (the mixed-language condition). Subjects were asked to remember

either the gist of the sentence or both the gist and the language in which the sentence was written. Tsao found that linguistic integration occurs across different languages. He also found that subjects could discriminate between old and new sentences in the single-language condition and between old and translated old sentences in the mixed-language conditions. So, he suggested that some information about language and about what idea occurred in what language was retained.

In his second experiment, Tsao employed Kintsch and Monk's (1972) paradigm to study the storage of sentence information presented in different languages. Again, Chinese-English bilinguals were the subjects. The results showed that it took longer for subjects to read mixed-language paragraphs than the single-language paragraphs. However, after the subjects comprehended the paragraph, the reaction time for answering inferential questions concerning the contents of the paragraph they had just read was the same for both mixed-language and single-language conditions. In other words, after the sentences are comprehended and the semantic contents are stored away in a core code or system, subjects have free access to this information and can convert the information into any form of language in which they are required to respond. Tsao concluded that the underlying representation of information from connected discourse is propositional; verbatim details may well be retained but they do not influence the process of reasoning and decision-making.

In sum, from both sentence verification and sentence integration experiments, we may conclude that higher level processing is not affected by variations in orthographies.

SUMMARY AND CONCLUSION

There is an inseparable relation between written language and spoken language--they both are essential communication tools in human societies and to some extent the former is parasitic on the latter. There are many writing systems for many different languages. Essentially, they can be divided into three categories based upon their various grapheme-meaning relations: logographic, syllabic, and alphabetic. We have reviewed most of the major experimental work done with these different types of orthographies and have compared the similarities and differences between them in terms of a visual information processing framework. We have found that indeed in lower level processing, different orthographic symbols were processed differently in terms of visual scanning, perceptual demands, involvement of different pathways between print and meaning, and cerebral lateralization functions. However, when we consider visual information processing at the higher levels, we find no difference with respect to word recognition, working memory strategies, inferences, and comprehension. This evidence suggests that reading is a universal phenomenon, a culture-free cognitive activity, once people in different language systems have acquired the ability to decipher the written symbols. Thus, Gibson and Levin (1975) aptly describe the state of affairs as follows:

> The findings do not mean that the process of reading is not influenced by the nature of the different writing systems, but that the outcomes are alike. It seems reasonable that different writing

> systems which relate to language at different levels will involve attention to and abstraction of different aspects of the orthographic system. Readers of a syllabary must search for invariances at one level, readers of an alphabetic system, at another level. But the skilled readers of one system are able to read as efficiently as skilled readers of another. (p. 165).

These statements have been supported by the present review, which has indicated that while reading behavior at the macro level seems not to be affected by orthographic variations, the information processing strategies at the perceptual level are affected by how meaning is represented in the printed symbols. Given such differences in the bottom-up processes required in transforming the visual-spatial arrays into meaning units, beginning readers of different writing systems apparently face different learning tasks when they are taught how to decipher printed symbols. The match between the task demands imposed by various writing systems and the developing cognitive structure of the beginning reader is an essential factor for success in such learning.

The three major writing systems reviewed assume three different types of script-speech relations. Chinese logography represents speech at the level of the morpheme rather the word, so that each logogram stands for the smallest type of meaningful unit and hence its form remains constant regardless of syntactic structure. That is, grammatical marking elements, such as tense, plural, gender, and so on, are introduced by adding other morpheme characters rather than by modifying the form of a particular character. For example, in Chinese logographs, go, went, and gone are expressed by exactly the same character (去) and both ox and oxen are expressed by the single character (牛). This perceptual constancy provides a certain advantage over those writing systems, such as the English alphabet, that require the marking of grammatical inflections at the word level. Thus, a reader learning a logographic system may have initial success as long as the characters to be learned are distinctively different; but as more characters are introduced, there are bound to be similarities to the previously learned characters (after all, the number of basic strokes in Chinese character formation is only eight). Then, whatever cues the young reader was using tend to fail, confusion sets in, and learning is disrupted until other memory strategies are acquired (Samuels, 1976).

The syllabary represents speech at the level of the syllable, a much more easily segmental unit than the phoneme, with a reduced set of symbols. For a beginning reader, the match between symbol and perceived sound segment makes the translation of visual arrays into speech code an easy task. The concept of mapping the secondary linguistic activity (reading) onto the primary linguistic activity (speech) can be acquired earlier through direct perceptual-associative links. However, the initial success of learning a syllabary starts to collapse as soon as more lexical items are learned and the problem of homophones sets in, and confusions over segmentation (examples in English would be to-gether vs. to-get-her; a-muse vs. am-use) pile up during ordinary reading (Suzuki, 1963). Special processing strategies are required, with great demands on the reader for the linguistic parsing of a syllabary text (Scribner & Cole, 1978).

Finally, an alphabetic writing system represents speech at the morphophonemic level such that the grapheme-sound-meaning relation is more or less opaque, requiring a more analytical processing strategy to unpack the meaning encoded in words, which are composed of a further reduced set of symbols. The abstractness of such a multilevel representation may be optimal for fluent readers (Chomsky & Halle, 1968), but it poses a great deal of difficulty for those beginning readers whose cognitive ability has not achieved the level necessary for extracting the orthographic regularities embedded in the written words. Liberman, Shankweiler, Liberman, Fowler, and Fischer (1977) reported a high correlation between children's reading ability and phoneme segmentation performance. They carried out a longitudinal study with nursery-school, kindergarten, and first-grade children and found that when children of all ages were asked to identify the number of phonetic segments in spoken utterances, none of the 4-year olds could segment by phoneme whereas nearly half (46%) could segment by syllable. At age 6, 70% succeeded in phoneme segmentation while 90% were successful in syllable segmentation. They then tested the same children at the beginning of the second school year and found that half of the children in the lowest third of the class on a reading achievement test had failed the phoneme segmentation task the previous June. On the other hand, all the children who passed the phoneme segmentation task scored in the top third on the reading achievement test. They concluded that the ability to break down the spoken utterance into its components is crucial to reading acquisition. Mattingly (1972) proposed that development of competence in reading requires that the internal structure of one's language be made explicit. "Linguistic awareness" refers to the individual's conscious knowledge of the types and levels of linguistic structures that characterize the spoken utterance. A beginning reader has to know the spelling-to-sound rules of English in order to recognize an old word, and a mature reader uses these rules to assign a pronunciation to a printed word that he has not seen before. The critical role of Mattingly's "linguistic awareness" in learning to read has been supported by several recent reading studies in English, which has a phonologically deep orthographic structure (Liberman et al., 1977; Liberman & Shankweiler, 1979), and in Serbo-Croatian, which has a phonologically shallow orthography (Lukatela & Turvey, 1980).

A critical question that deals directly with the relation between orthography and reading should be raised at this point: What aspects of sentences in spoken language do different orthographies attempt to transcribe? The traditional classification of orthographies into logographic, syllabary, and alphabetic modes seems to imply that each mode transcribes sentences in radically different ways (but see Mattingly, Note 7). However, from our review of the literature, the generalization seems to be that all orthographies attempt to transcribe sentences at the level of words and, furthermore, the transcription of words is morphemic in nature. This point seems unnecessarily obvious in Chinese logography. The morphophonemic character of an alphabetic orthography is also obvious in the case of a language with a relatively "deep" phonology, such as English or French. An example of such representation can be seen in the transcription of the words heal, health, healthy (Chomsky & Halle, 1968). Mattingly (Note 7) has convincingly demonstrated that the same morphophonemic principle holds for orthographies with shallow phonology, such as Vietnamese and Serbo-Croatian, as well as for syllabary orthography, such as Japanese. This characterization of orthography suggests that in the actual process of reading, the analysis of a sentence

begins with its lexical content and not with its phonetic representation, since neither Chinese nor English transcribes words in phonetic forms. In fact, in sentence processing, regardless of the type of orthography, phonetic representation is used for the purpose of refreshing the information in short-term memory, especially when the material is difficult (Hardyck & Petrinovich, 1970; Tzeng et al., 1977). This conceptualization is consistent with the observation that differences due to orthographic variation in the visual processing of print occur only before but not after word recognition.

Given this argument that all orthographies attempt to transcribe sentences at the word level, the next question is whether different ways of achieving such a transcription also create different pathways between print and the lexicon. The answer is positive and at least two pathways can be readily identified. The phonologically based route represents a procedure or rule learning of knowing how and the visually based route represents an associative learning of knowing that. These two types of knowledge may have different neurological realizations (Cohen & Squire, 1980). In principle, different dyslexic patterns (Marshall & Newcombe, 1973) may result from the selective impairment of these two pathways or their combinations. However, experimental data together with clinical observations are very much needed to support all these arguments. Aphasic studies across different orthographies would certainly reveal important details about these different pathways.

Another question that needs to be answered is whether there is an optimal orthography for the purpose of reading. Anthropologists are generally sensitive to such a question, since it may imply a linguistic chauvinism--the belief that one's own orthography is the best of all possible orthographies. But the arguments advanced about written languages should be carefully distinguished from those concerning spoken language. In speech, moving our tongues and maneuvering air through our supralaryngeal tracts are no more foreign to us than programming our arms to move, wave, grasp, or make gestures. Using written languages, on the other hand, requires the utilization of something external to us: conventional notational systems invented by human beings. Changes in spoken language follow a more or less universal principle of biological evolution whereas maintenance or change in written languages is usually by sociocultural and cognitive factors, which may sometimes be as arbitrary as a dictator's decision. The apparent heterogeneity of orthographies may also imply inequality in the ease of achieving reading efficiency. It is therefore legitimate as well as important to raise the question about criteria of an optimal orthography, with or without respect to different spoken languages. No answer can be provided here. However, clues for a plausible answer may be obtained in Wang (in press).

One thing is sure: We cannot study a writing system without also considering the spoken language it attempts to transcribe. From history we learn that the development of a particular writing system is always constrained by the linguistic properties of its corresponding spoken language. The fact that the Chinese writing system adopts a logographic system and stops at the morphosyllabic level reflects the monosyllabic nature of its morphemes and the lack of morphological inflection. When the Japanese borrowed Chinese characters to transcribe their spoken language, additional symbols were required to represent grammatical inflections. Hence, Japanese scholars of those early days had to take some Chinese characters apart and derive from

them the sound symbols, namely, the kana syllable elements (Wang, in press). But due to the simplicity of the syllabic structure and the limited number of syllables in their spoken language (no more than 90 different syllables are used; hence, the problem of homophones), the Japanese adopted both syllabary and logographic scripts. For Koreans, who also borrowed Chinese characters to transcribe their spoken language, the writing system had to go one more step to the level of alphabet in order to meet the perceptual demands imposed by its much richer syllable structure (Martin, 1972).

From these examples, one can see that the relation between script and speech in any language exhibits a principle of mutual compatibility. That is, the relation suggests that through writing, properties of substance (meaning) and surface (script) enter into invariant combinations (at the level of words) to comprise a speech-relevant description of the semantic intents. In other words, when we read an array of graphemic symbols, we not only register the physical properties (shape, length, width, space, etc.) of the print, but also perceive the unique, abstract properties of speech that are afforded (supported or furnished) by this particular type of script. Such a complementarity of the script and the speech is best captured by the notion of affordance proposed by Gibson (1977). Thus, in this sense, no writing system should be claimed to be more advanced than others. The principle of mutual compatibility also implies that successful reading depends on the maturation and the awareness of one's own spoken language (Mattingly, 1972).

Man stands alone in history as the sole creature on earth who invents written symbols and who also benefits from these symbols. Since these new symbols are to some extent arbitrary inventions external to our organismic structure, both accommodation and assimilation processes must have worked at their extremes in order for us to achieve efficiency in manipulating them. It took a span of many thousand years for our ancestors to come up with a system that works for a particular language and it takes a great deal of effort on the part of a modern learner to become a fluent reader. The diversity of writing systems provides excellent opportunities for investigators of human cognition to examine how children of different languages adjust themselves to meet various task demands imposed by different orthographies. Once we understand something about the kind of advantage or disadvantage that a certain type of orthographic representation can bestow, we would be in a better position to understand how man can come to invent them. Once we are able to understand the script-speech relations in various writing systems and find out effects of such orthographic variations on our reading behaviors, we would be in a better position to "unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history" (Huey 1908/1968, p. 6).

REFERENCE NOTES

1. Elman, J. L. Asymmetries for the categorization of kanji nouns, adjectives, and verbs presented to the left and right visual fields. Linguistic Notes from La Jolla, No. 7, 1979.
2. Shimamura, A., & Hunt, E. Stroop interference tests with kanji and kana scripts. Unpublished paper, University of Washington, Seattle, 1978.

3. Smith, M. C. Are Chinese ideographs processed like pictures? Paper presented at the Annual Meeting of the Psychonomic Society, St. Louis, November, 1980.
4. Meyer, D. E., & Ruddy, M. G. Lexical memory retrieval based on graphemic and phonemic representations of printed words. Paper presented at the Annual Meetings of the Psychonomic Society, St. Louis, 1973.
5. Tzeng, O. J. L., Malley, K. W., Hung, D. L., & Dreher, J. How do you draw a coffee cup? Paper presented at the Annual Meeting of the Psychonomic Society, St. Louis, November, 1980.
6. So, K. F., Potter, M. C., & Friedman, R. Reading in Chinese and English: Naming versus understanding. Unpublished paper, Massachusetts Institute of Technology, 1977.
7. Mattingly, I. G. Reading, linguistic awareness and language acquisition. Paper presented at the International Reading Association Seminar on Linguistic Awareness and Learning to Read, Victoria, B.C., June, 1979.

REFERENCES

Asayama, T. Über die Aphasie bei Japanern. Deut. Arch. Klin. Med., 1914, 113, 523-529.

Baddeley, A. D., & Hitch, C. Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8), 1974, 47-90.

Bannatyne, A. Language, reading and learning disabilities. Springfield, Ill.: Thomas, 1976.

Baron, J. Phonemic stage not necessary for reading. Quarterly Journal of Experimental Psychology, 1973, 25, 241-246.

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 386-393.

Baron, J., & Treiman, R. Use of orthography in reading and learning to read. In R. L. Venezky & J. F. Kavanagh (Eds.), Cross language research in reading, orthography, and dyslexia. Baltimore, Md.: University Park Press, 1980.

Barron, R. N., & Baron, J. How children get meaning from printed words. Child Development, 1977, 48, 487-594.

Bertelson, P., & Tisseyre, F. Set and lateral asymmetry in the perceived sequence of speech and non-speech. In P. Rabbitt & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Besner, D., & Coltheart, M. Ideographic and alphabetic processing in skilled reading of English. Neuropsychologia, 1979, 17, 467-472.

Biederman, I., & Tsao, Y. C. On processing Chinese ideographs and English words: Some implications from Stroop-test results. Cognitive Psychology, 1979, 11, 125-132.

Bower, T. G. R. Reading by eye. In H. Levin & J. Williams (Eds.), Basic studies in reading. New York: Basic Books, 1970.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Broadbent, D. E. Word frequency effect and response bias. Psychological Review, 1967, 74, 1-15.

Broca, P. Remarques sur le siege de la faculte du langage articule suivies d'une observation d'aphemie. Bulletin de la Societe Anatomique de Paris, 1861, 6, 330-357.

Brooks, L. R. Visual pattern in fluent word identification. In A. S. Reber & D. Scarborough (Eds.), Toward a psychology of reading. Hillsdale, N.J.: Erlbaum, 1977.

Campbell, D. T., & Stanley, J. C. Experimental and quasiexperimental designs for research. Chicago: Rand McNally, 1963.

Carpenter, P. A., & Just, M. A. Sentence comprehension: A psycholinguistic processing model of verification. Psychological Review, 1975, 82, 45-73.

Cattell, J. M. The time taken up by cerebral operations. Mind, 1886, 11, 220-242.

Chang, C. Y. A study of the relative merits of the vertical and horizontal lines in reading Chinese print. Archives of Psychology, 1942, 276, Columbia University.

Chen, L. K., & Carr, H. The ability of Chinese students to read in vertical and horizontal direction. Journal of Experimental Psychology, 1926, 9, 110.

Chomsky, C. Reading, writing and phonology. Harvard Educational Review, 1970, 40, 287-309.

Chomsky, N., & Halle, M. The sound pattern of English. New York: Harper and Row, 1968.

Chou, S. K. Reading and legibility of Chinese characters. Journal of Experimental Psychology, 1929, 12, 156.

Chu-Chang, M., & Loritz, D. J. Phonological encoding of Chinese ideographs in short-term memory. Language Learning, 1977, 27, 344-352.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Cohen, N. J., & Squire, L. R. Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science, 1980, 210, 207-210.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

Corballis, M. C., & Beale, I. L. On telling left from right. Scientific American, 1971, March, 96.

Corcoran, D. W. Acoustic factors in proof reading. Nature, 1967, 214, 851.

Corcoran, D. W., & Weening, D. L. Acoustic factors in visual search. Quarterly Journal of Experimental Psychology, 1968, 20, 83-85.

Cosky, M. J. Word length effects in word recognition. Unpublished doctoral dissertation, University of Texas, Austin, 1975.

Critchley, M. Aphasia in polyglots and bilinguals. Brain and Language, 1974, 1, 15-27.

Crowder, R. G. The sound of vowels and consonants in immediate memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 587-596.

Cutting, J. E. Two left-hemisphere mechanisms in speech perception. Perception & Psychophysics, 1974, 1, 601-612.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. Phonological recoding and lexical access. Memory & Cognition, 1978, 6, 391-402.

de Agostini, M. A propos de l'agraphie des aphasiques sensoriels: Etude comparative Italien-Francais. Language, 1977, 47, 120-130.

Dreyfuss, H., & Fuller, R. (Eds.). Symbol sourcebook: An authoritative guide to international graphic symbols. New York: McGraw Hill, 1972.

Dyer, F. N. Color-naming interference in monolinguals and bilinguals. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 297-302.

Elkind, D., & Weiss, J. Studies in perceptual development III: Perceptual exploration. Child Development, 1967, 38, 1153-1161.

Erickson, D., Mattingly, I., & Turvey, M. Phonetic activity and reading: An experiment with kanji. Language and Speech, 1977, 20, 384-403.

Fang, S. P., Tzeng, O. J. L., & Alva, L. Intra- versus inter-language Stroop effects in two types of writing systems. Memory & Cognition, in press. (Also Haskins Laboratories Status Report on Speech Research, 1981, this volume.)

Feldman, L. B., & Turvey, M. Words written in Kana are named faster than the same words written in Kanji. Language and Speech, 1980, 23, 141-147.

Ferguson, N. Pictographs and prereading skills. Child Development, 1975, 46, 786-789.

Fodor, J., & Bever, T. G. The psychological reality of linguistic segments. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 414-420.

Frith, U. Cognitive processes in spelling. New York: Academic Press, 1980.

Gelb, I. J. A study of writing. Chicago: University of Chicago Press, 1952.

Geschwind, N. The organization of language and the brain. Science, 1970, 170, 940-944.

Geschwind, N. Dyslexia. Science, 1971, 173, 190.

Gibson, E. J. Trends in perceptual development: Implications for reading process. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Del.: International Reading Association, 1976.

Gibson, E. J. How perception really develops: A view from outside the network. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Gibson, E. J., & Levin, H. The psychology of reading. Cambridge, Mass.: MIT Press, 1975.

Gibson, J. J. The theory of affordances. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing: Towards an ecological psychology. Hillsdale, N.J.: Erlbaum, 1977.

Gleitman, L. R., & Rozin, P. The structure and acquisition of reading: Relation between orthography and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY conferences. Hillsdale, N.J.: Erlbaum, 1977.

Goodnow, J., Friedman, S. L., Bernbaum, M., & Lehman, E. B. Direction and sequence in copying: The effect of learning to write in English and Hebrew. Journal of Cross-Cultural Psychology, 1973, 4, 263-282.

Hardyck, C., & Petrinovich, L. F. Subvocal speech and comprehension level of reading material. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 647-652.

Hardyck, C., & Petrinovich, L. F. Left-handedness. Psychological Bulletin, 1977, 84, 385-404.

Hardyck, C., Tzeng, O. J. L., & Wang, W. S-Y. Cerebral lateralization effects in visual half-field experiments. Nature, 1977, 269, 705-707.

Hardyck, C., Tzeng, O. J. L., & Wang, W. S-Y. Cerebral lateralization of function and bilingual decision processes: Is thinking lateralized? Brain and Languge, 1978, 5, 56-71.

Hatta, T. Asynchrony of lateral onset as a factor in difference in visual fields. Perceptual and Motor Skills, 1976, 42, 163-166.

Hatta, T. Recognition of Japanese Kanji in the left and right visual field. Neuropsychologia, 1977, 15, 685-688.

Hecaen, H., & Albert, M. Human neuropsychology. New York: John Wiley & Son, 1978.

Heron, W. Perception as a function of retinal locus and attention. American Journal of Psychology, 1957, 70, 38.

Hinshelwood, J. Congenital word-blindness. London: Lewis, 1917.

Hirata, K., & Osaka, R. Tachistoscopic recognition of Japanese letter materials in left and right visual fields. Psychologia, 1967, 10, 17-18.

Huang, J. T., & Liu, I. M. Paired-associate learning proficiency as a function of frequency count, meaningfulness, and imagery value in Chinese two-character ideograms. Acta Psychologica Taiwanica, 1978, 20, 1-14.

Huey, E. B. The psychology and pedagogy of reading. Cambridge, Mass.: MIT Press, 1968. (Originally published, 1908)

Hung, D. L., Tzeng, O. J. L., & Warren, D. H. A chronometric study of sentence processing in deaf children. Cognitive Psychology, in press.

Huttenlocher, J. Encoding information in sign language. In J. F. Kavanagh & J. E. Cutting (Eds.), The role of speech in language. Cambridge, Mass.: MIT Press, 1975.

Imura, T. Aphasia: Characteristic symptoms in Japanese. Journal of Psychiatric Neurology, 1943, 47, 196-218.

Jensen, A. R. Educability and group differences. New York: Harper and Row, 1973.

Johnston, J. C. Word perception: Clues from studying the word-superiority effect. In O. J. L. Tzeng & H. Singer (Eds.), Perception of print: Reading research in experimental psychology. Hillsdale, N.J.: Erlbaum, in press.

Just, M. A., & Carpenter, P. A. Comparative studies of comprehension: An investigation of Chinese, Norwegian, and English. Memory & Cognition, 1975, 3, 465-473.

Kavanagh, J. F., & Venezky, R. L. (Eds.). Orthography, reading, and dyslexia. Baltimore, Md.: University Park Press, 1980.

Keele, S. W. Attention demands of memory retrieval. Journal of Experimental Psychology, 1972, 93, 245-248.

Kimura, D. The asymmetry of the human brain. Scientific American, 1973, 228, 70-78.

Kimura, D. The neural basis of language qua gesture. In H. Whitaker & H. A. Whitaker (Eds.), Studies in neurolinguistics (Vol. 2). New York: Academic Press, 1976.

Kintsch, W., & Buschke, H. Homophones and synonyms in short-term memory. Journal of Experimental Psychology, 1969, 80, 403-407.

Kintsch, W., & Monk, D. Storage of complex information in memory: Some implication of the speed with which inferences can be made. Journal of Experimental Psychology, 1972, 94, 25-32.

Kintsch, W., & Van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review 1978, 85, 363-416.

Kleiman, G. M. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-339.

Klima, E. S., & Bellugi, U. The signs of language. Cambridge, Mass.: Harvard University Press, 1979.

Kolers, P. A. Reading and knowing. Canadian Journal of Psychology, 1979, 33, 106-117.

Kolers, P. A., Wrolstad, M. E., & Bouma, H. Processing of visible language. New York: Plenum Press, 1979.

Koriat, A., & Levy, I. Figural symbolism in Chinese ideographs. Journal of Psycholinguistic Research, 1979, 8, 353-365.

Kroeber, A. L. Anthropology: Race, language, culture, psychology, prehistory. New York: Harcourt Brace, 1948.

Kugelmass, S., & Lieblich, A. Impact of learning to read on directionality in perception: A further cross-cultural analysis. Human Development, 1979, 22, 406-415.

Laboratory of Comparative Human Cognition. What's cultural about cross-cultural cognitive psychology? Annual Review of Psychology, 1979, 30, 145-172.

Lenneberg, E. H. Biological foundations of language. New York: Wiley, 1967.

Levy-Schoen, A., & O'Regan, K. The control of eye movements in reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), Processing of visible language. New York: Plenum Press, 1979.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, Md.: University Park Press, 1980.

Liberman, I. Y., & Shankweiler, D. Speech, the alphabet, and teaching to read. In L. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, N.J.: Erlbaum, 1979.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY conferences. Hillsdale, N.J.: Erlbaum, 1977.

Liu, I. M., & Chuang, C. J. Scaling 1200 Chinese characters for meaningfulness. Acta Psychologica Taiwanica, 1970, 12, 33-52.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Lukatela, G., & Turvey, M. T. Some experiments on Roman and Cyrillic alphabets. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, Md.: University Park Press, 1980.

Luria, A. R. Differences between disturbance of speech and writing in Russian and French. International Journal of Slavic Linguistics and Poetics, 1960, 3, 13-22.

Lyman, R., Kwan, S. T., & Chao, W. H. Left-occipital-parietal brain tumor with observations on alexia and agraphia in Chinese and English. Chinese Medical Journal, 1938, 54, 491-516.

MacKay, D. G. Input testing in the detection of misspellings. American Journal of Psychology, 1972, 85, 121-127.

Marshall, J. C., & Newcombe, F. Patterns of paralexia: A psycholinguistic approach. Journal of Psycholinguistic Research, 1973, Vol.2, 3, 175-199.

Martin, S. E. Nonalphabetic writing systems. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationship between speech and reading. Cambridge, Mass.: M.I.T. Press, 1972.

Massaro, D. M. Understanding language: An information processing model of speech perception, reading, and psycholinguistics. New York: Academic Press, 1975.

Mattingly, I. G. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationship between speech and reading. Cambridge, Mass: MIT Press, 1972.

McNeill, D. The acquisition of language: The study of developmental psycholinguistics. New York: Harper and Row, 1970.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Functions of graphemic and phonemic codes in visual word recognition. Memory & Cognition, 1974, 2, 309-321.

Meyer, D. E., & Schvaneveldt, R. W. Meaning, memory structure, and mental processes. In C. N. Cofer (Ed.), The structure of human memory. San Francisco, Ca.: W. H. Freeman and Co., 1976.

Mishkin, M., & Forgays, G. D. Word recognition as a function of retrieval locus. Journal of Experimental Psychology, 1952, 43, 43-48.

Miura, T. The word superiority effect in a case of Hiragana letter string. Perception & Psychophysics, 1978, 24, 505-508.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Nelson, T. M., & Ladar, C. J. Judgment of meaningfulness of Chinese characters by English speaking observers. Visible Language, 1976, 10, 129-143.

Orbach, J. Differential recognition of Hebrew and English words in right and left visual fields as a function of cerebral dominance and reading habits. Neuropsychologia, 1966, 5, 127-134.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart, and Winston, 1971.

Patterson, K., & Bradshaw, J. Differential hemisphere mediation of nonverbal visual stimuli. Journal of Experimental Psychology: Human Perception and Performance, 1975, Vol.1, 246-252.

Peuser, G., & Leischner, A. Impairments of phonetic script in an aphasic. (1974) In M. Paradis (Ed.), Readings on aphasia in bilinguals and polyglots. Montreal: Didier, 1980, in press.

Poizner, H., Battison, R., & Lane, H. Cerebral asymmetry for American Sign Language: The effect of moving stimuli. Brain and Language, 1979, 7, 351-362.

Potter, M. C., & Faulconer, B. A. Time to understand pictures and words. Nature, 1975, 253, 437-438.

Potter, M. C., Kroll, J. F., & Harris, C. Comprehension and memory in rapid sequential reading. In R. Nickerson (Ed.), Attention and performance (VIII). Hillsdale, N.J.: Erlbaum, 1980.

Preston, M. S., & Lambert, W. E. Interlingual interference in a bilingual version of the Stroop color-word task. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 295-301.

Pynte, J. Readiness for pronunciation during the reading process. Perception & Psychophysics, 1974, 16, 110-112.

Rayner, K. Eye movement in reading and information processing. Psychological Bulletin, 1978, 85, 618-660.

Reicher, G. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 276-280.

Rohrman, N. L., & Gough, P. Forewarning, meaning and semantic decision latency. Psychonomic Science, 1967, 9, 217-218.

Rosinski, R. R., Gollinkoff, R. M., & Kukish, K. C. Automatic semantic processing in picture-word interference task. Child Development, 1975, 46, 247-253.

Rozin, P., Poritsky, S., & Sotsky, R. American children with reading problems can easily learn to read English represented by Chinese characters. Science, 1971, 171, 1264-1267.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Sabol, M. A., & DeRosa, D. V. Semantic encoding of isolated words. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 58-68.

Sakamoto, T., & Makita, K. Japan. In J. Downing (Ed.), Comparative reading: Cross national studies of behavior and processes in reading and writing. New York: Macmillan, 1973.

Salapatek, P. Visual scanning of geometric figures by the human newborn. Journal of Comparative and Physiological Psychology, 1968, 66, 247-258.

Samuels, S. J. Modes of word recognition. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Del.: International Reading Association, 1976.

Sasanuma, S. An analysis of writing errors in Japanese aphasic patients: Kanji vs. Kana words. Cortex, 1972, 8, 256-282.

Sasanuma, S. Kanji vs. Kana processing in alexia with transient agraphia: A case report. Cortex, 1974, 10, 89-97. (a)

Sasanuma, S. Impairment of written language in Japanese aphasics Kana vs. Kanji processing. Journal of Chinese Linguistics, 1974, 2, 141-157. (b)

Sasanuma, S. Kana and Kanji processing in Japanese aphasics. Brain and Language, 1974, 2, 369-383. (c)

Sasanuma, S., & Fujimura, O. Selective impairment of phonetic and nonphonetic transcription of words in Japanese aphasic patients: Kana vs. Kanji in visual recognition and writing. Cortex, 1971, 7, 1-18.

Sasanuma, S., Itoh, M., Mori, K., & Kobayashi, Y. Tachistoscopic recognition of Kana and Kanji words. Neuropsychologia, 1977, 15, 547-553.

Scribner, S., & Cole, M. Unpacking literacy. Social Science Information, 1978, 17, 19-40.

Seidenberg, M., & Tanenhaus, M. K. Orthographic effects on rhyme monitoring. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 546-554.

Shen, E. An analysis of eye movements in the reading of Chinese. Journal of Experimental Psychology, 1927, 10, 158-183.

Singer, H. The substrata-factor theory of reading: Theoretical design for teaching reading. In J. A. Figurel (Ed.), Challenge and experiment in reading: Proceedings of the Seventh Annual Conference of the International Reading Association. New York: Scholastic Magazines, 1962, 7, 226-232.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition threshold for words. Journal of Experimental Psychology, 1952, 43, 195-201.

Sperry, R. W., Gazzaniga, M. S., & Bogen, J. E. Interhemispheric relationships: The neocortical commissures; syndromes of hemispheric disconnection. In P. J. Vinken & G. W. Bruya (Eds.), Handbook of clinical neurology (Vol. 4). Amsterdam: North-Holland, 1969.

Spoehr, K. T., & Smith, E. E. The role of orthographic and phonotactic rules in perceiving letter patterns. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 21-34.

Sternberg, S. The discovery of processing stages: Extension of Donder's method. Acta Psychologica, 1969, 30, 276-315.

Stroop, J. R. Studies of interference in serial verbal reaction. Journal of Experimental Psychology, 1935, 18, 643-662.

Suzuki, T. A semantic analysis of present day Japanese. Tokyo: Keio Institute of Cultural and Linguistic Studies, 1963.

Treiman, R., Baron, J., & Luk, K. Type of orthography affects use of sound in silent reading. Journal of Chinese Linguistics, 1981, 9, 116-125.

Tsao, Y. C. Comprehension processes in bilinguals. Unpublished doctoral dissertation, State University of New York, Buffalo, 1973.

T·:, T. C. The effects of different arrangements of Chinese language upon speed and comprehension of silent reading. Journal of Genetic Psychology, 1930, Dec, 321.

Tzeng, O. J. L., & Hung, D. L. Reading in the nonalphabetic writing system: Some experimental studies. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, Md.: University Park Press, 1980.

Tzeng, O. J. L., & Hung, D. L. Linguistic determinism: A written language perspective. In O. J. L. Tzeng & H. Singer (Eds.), Perception of print: Reading research in experimental psychology. Hillsdale, N.J.: Erlbaum, in press.

Tzeng, O. J. L., Hung, D. L., Cotton, B., & Wang, W. S-Y. Visual lateralization effect in reading Chinese characters. Nature, 1979, 282, 499-501.

Tzeng, O. J. L., Hung, D. L., & Garro, L. Reading the Chinese characters: An information processing view. Journal of Chinese Linguistics, 1978, 6, 287-305.

Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.

Tzeng, O. J. L., & Singer, H. (Eds.), Perception of print: Reading research in experimental psychology. Hillsdale, N.J.: Erlbaum, in press.

Vaid, J., & Genesee, F. Neuropsychological approaches to bilingualism: A critical review. Canadian Journal of Psychology, in press.

Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.

Venezky, R. L. Overview: From Sumer to Leipzig to Bethesda. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, Md.: University Park Press, 1980.

Wang, W. S-Y. The Chinese language. Scientific American, 1973, 228, 50-60.

Wang, W. S-Y. Language structure and optimal orthography. In O. J. L. Tzeng & H. Singer (Eds.). Perception of print: Reading research in experimental psychology. Hillsdale, N.J.: Erlbaum, in press.

Warren, R. Stimulus encoding and memory. Journal of Experimental Psychology, 1972, 94, 90-100.

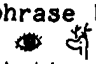
Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. Journal of Experimental Psychology, 1965, 70, 102-108.

Wood, C. C., Goff, W. R., & Day, R. S. Auditory evoked potentials during speech perception. Science, 1971, 173, 1248-1251.

Yeh, J. S., & Liu, I. M. Factors affecting recognition thresholds of Chinese characters. Acta Psychologica Taiwanica, 1972, 14, 113-116.

FOOTNOTES

[1]Examples of such ideograms during the early development of written scripts can be found in many different parts of the world. Huey (1908/1968), in his monumental book on reading, gives many excellent examples to illustrate the principle of metonymy. For those researchers who are interested in the issue of metaphor, these ancient ideograms and the rules behind their formations can be a very useful resource for discovering how people use metaphors.

[2]A representation of word or phrase by pictures that suggest how it is said in the spoken language, e.g.,  for idea. The rebus system is a hybrid of picture and sound representations.

[3]There is, however, some experimental evidence suggesting that the rate of reading English may be limited by the reader's horizontal eye movements. With a method of RSVP (Rapid serial visual presentation), Potter, Kroll, and Harris (1980) demonstrate that when eye movements are not required, readers are able to comprehend text presented as rapidly as 12 wps (word per second), more than twice as fast as people normally read. Interestingly, reading in a RSVP manner is highly similar to the way a Chinese reader reads a vertically arranged text. Results of these RSVP studies suggest that there may be some yet-to-be-discovered advantages of the Chinese way, after all.

[4]The term lateralization refers to the specialization of the left and right hemispheres of the brain for different functions. The rationale behind the visual hemi-field experiment and the actual experimental set-up will be discussed in a later section.

[5]Strictly speaking, the proposition that Chinese characters do not specify sounds of the spoken language is not correct. We have already noted phonograms (see Figure 1) constitute a majority of modern day Chinese logograms.

VISUAL WORD RECOGNITION IN SERBO-CROATIAN IS NECESSARILY PHONOLOGICAL

Laurie Beth Feldman

Abstract. In a naming task conducted with bi-alphabetic readers of Serbo-Croatian, it was shown that letter strings that can be assigned both a Roman and a Cyrillic alphabet reading incur longer latencies than the unique alphabet transcription of the same word, and that the magnitude of the difference depended on the number of ambiguous characters in the ambiguous letter string. While this within-word phonological ambiguity effect obtained for both words and pseudowords, it was more consistent with words. The same pattern of results occurred in a lexical decision task, and the correlation between latencies (for words and pseudowords) in the two tasks was significant. It was concluded that both lexical decision and naming in Serbo-Croatian necessarily involve a phonological strategy.

INTRODUCTION

Alphabetic Writing Systems: The Legacy of a Phonographic Orthography

Writing systems differ in terms of the units with which they transcribe the spoken language. Logographies such as Chinese and Japanese Kanji have characters that correspond to words or morphemes. Japanese Kana and Hebrew are examples of (approximately) syllabic orthographies where each character of the written language corresponds most closely to a syllable unit (Gelb, 1952). Perhaps the most complex orthographies to learn are alphabetic, where words are transcribed by phonemes that are abstract units relative to the syllable and the word (Mattingly, 1972). Both the syllabary and the alphabet are phonographic orthographies where the characters that comprise the written form correspond most closely to segments of speech. In the evolution of writing systems based on the spoken language, the introduction of a phonographic principle represents greater complexity as it exploits the abstract relation between orthographic characters that comprise the word and the word as spoken.

Acknowledgment. Special thanks to Michael Turvey, Aleksandar Kostić, Georgije Lukatela, and Carol Fowler. Thanks also to Jacqueline Sachs, Manolo Gurjanov, Milan Savić, Leonard Katz, Vicki Hanson, Predrag Ognjenović, Bojan Gligorijević, and Borislav Lorenc. This work was part of a doctoral dissertation presented to the University of Connecticut.

The research was supported in part by NICHD Grant 08495 to the University of Belgrade, in part by NICHD Grant HD 01994 to Haskins Laboratories, and in part by NSF dissertation fellowship BNS 7924409 to Laurie Feldman. In addition, funds were contributed by the Faculties of Philosophy and Electrical Engineering at the University of Belgrade and by the University of Connecticut.

Consequently, this suggests greater demands on the analytic capabilities of the reader. The benefits, however, would appear to compensate the disadvantages: As far as mastering a written vocabulary, the phonographic principle reduces the task of learning and recognizing new word forms (Gibson & Levin, 1975; Gleitman & Rozin, 1977).

Among alphabetic systems, the depth of the orthography and the relation between the written and spoken forms may vary. Written Serbo-Croatian respects a phonographic principle fully, retaining a very consistent relation between (classical) phoneme and grapheme. In contrast, the graphemes and phonemes in English are less direct and more variable in their mapping: English graphemes tend to represent (systematic) phonemes or morphophonemes. The consequence of this systematicity at the morphophonemic level is that for many words of English, the orthographic form does not directly specify the surface phonetic form. (For example, the morphological relationship of "HEAL" to "HEALTH" is captured in the written form of the words, while the specification of the differing vowel sounds is sacrificed.) In addition, the letter-sound correspondences are variable in English, as there are many exception words (e.g., "HAVE" versus "SAVE"). In general, theories of word recognition and reading have been described for English and have accommodated the idiosyncracies of the English orthography into an account of the strategies for word recognition. The present studies constitute an attempt to evaluate the word recognition strategies delineated for English when they are applied to the phonologically shallow orthography of Serbo-Croatian.

For alphabetic orthographies, a reader may derive a word's phonological form in one of three ways. Two of these may be termed both <u>phonological</u> and <u>word-nonspecific</u>, and one may be termed <u>visual</u> and <u>word-specific</u>. The two varieties of phonological word-nonspecific strategies are analytic in that they exploit the phonographic principle that relates the written form to the spoken form. Consequently, they can apply equally to both words and pseudowords and proceed independently of word-specific knowledge. Exploiting general grapheme-phoneme correspondence rules (Venezky, 1970) that abstractly map between print and speech is one possible strategy, and it will work successfully for any letter string that does not violate the correspondence rules. These mapping rules analyze independent grapheme units (Gough, 1972) or functional graphemes (Gibson, 1962, 1970) in order to arrive at a phonological description. Therefore, to the extent that the generation of a phonological code is the sole determiner of response time, recognition latencies for words and for pseudowords with similar orthographic structure should be equal, and latency should be a function of the number and complexity of independent graphemic units.

A second phonologically analytic, word-nonspecific strategy proposed minimizes the importance of individual grapheme-phoneme correspondences and promotes procedures involving the coordination and synthesis of several phonological representations (each of which may be a multi-letter unit). Here, the phonology of a letter string is derived by a process of (automatic) analogy based on its orthographic similarity to other strings of letters (Glushko, 1979, 1981). Pseudowords and words are pronounced by analogy with the same multi-letter units, termed orthographic neighborhood, as they occur in other real words rather than by application of context-insensitive letter-sound correspondence rules. In general, the two phonological, word-

nonspecific strategies subsume both words and pseudowords as they are analytic and, therefore, do not depend on the familiarity of particular lexical entries. To the extent that a phonological strategy is neutral with respect to lexical status of a letter string in a reading task, no interactions of lexicality with phonological variables are predicted. In general, evidence of phonological strategies is weak with real words that are exceptions to the grapheme-phoneme correspondence rules, that is, words such as "SWORD" or "TONGUE," but this may reflect how "regular" and "exception" words are defined (Glushko, 1979; Bauer & Stanovich, 1980). Nevertheless, for pseudowords and for words that are regular and obey the correspondence rules, either phonological strategy is always adequate.

The third strategy distinguishes among letter strings on the basis of their lexical status and the regularity of their letter-sound correspondences. This strategy is visual and word-specific (or morpheme specific, see Taft, 1979) and it entails a lexical look-up by which the reader goes from some aspect of the written form to an entry in the internal lexicon. Only in the lexicon is a phonological representation (as well as a phonetic representation) adequately specified for a particular word or sequence of morphemes. Within the lexicon, entries are organized and searched according to their frequency of occurrence and, within this strategy, response time is based on the ease of identifying a familiar visual form as an instance of a particular lexical entry. As a result, a strong correlation between reaction time and word frequency is usually interpreted as evidence of a lexical contribution to recognition (e.g., Rubenstein, Garfield, & Millikan, 1970). By a word-specific strategy, the essential part of the letter string is treated holistically, or at least not analytically in any phonographic sense. (In some accounts, e.g., Taft [1979], the letter string must be freed of affixes or nonessential segments. It is not always obvious how this procedure would operate given that the distinction between an essential and a nonessential letter sequence may require word-specific knowledge.) This strategy encompasses real words, both regular and exceptions, but it cannot apply to the reading of pseudowords, as a search of the lexicon would fail to locate an entry for them. To complement this strategy, one of the two word-nonspecific procedures need be introduced. This supplementary strategy is indistinguishable in kind from either of the phonologically-analytic word-nonspecific strategies described above, but since it is only used when the visual, word-specific strategy fails, it is only implemented for pseudowords.

In summary, in word recognition and reading, the phonologically analytic word-nonspecific strategies of grapheme-phoneme conversion or (automatic) analogy can be applied both for regular words and for pseudowords as they exploit a phonographic principle that is analytic and does not focus on particular lexical entries. The lexical strategy is not phonologically analytic. Because it is tied to a specific word's visual form, it can only succeed for real words. As the word-specific strategy is limited in effectiveness, it must be complemented sometimes by a phonological strategy. Whereas a word-specific strategy need not be sensitive to component orthographic structure or to phonological complexities, the effectiveness of a phonological strategy may depend on the lexical status of a letter string. There is empirical evidence that subjects have the option to alter the balance of recognition strategies according to the nature of the letter strings and the experimental task and that, at least in English, it is the relative

contribution of the phonological strategy that appears to vary (Coltheart, Besner, Jonasson, & Davelaar, 1979).

Evidence for a Phonological Recognition Strategy in English

In the literature on word recognition based on English, there are three sources of support for a phonological recognition strategy, although all are subject to frequent criticism: (1) effects of orthographic structure; (2) adherence to grapheme-phoneme correspondence rules; (3) effects of homophony. The nature of a strategy that exploits a phonographic principle implies the importance of orthographic structure to the processes of word recognition. In general, naming latency is sensitive to number of letters for both words and pseudowords, while in lexical decision, this structural variable is only important for pseudoword performance (Frederiksen & Kroll, 1976; Forster & Chambers, 1973). Likewise, the complexity and position of consonant clusters significantly affects naming but not lexical decision times (Frederiksen & Kroll, 1976). When naming protocols differ from lexical decision protocols, logical task requirements are generally invoked: where lexical decision requires specific word knowledge, naming may proceed independent of the lexicon (Baron, 1977; Coltheart et al., 1979). As a result, phonological effects demonstrated only in the naming task do not provide convincing evidence of a phonological recognition strategy.

With other factors controlled, time to decide that a letter string is a word (lexical decision) is generally shorter for those regular words that comply with grapheme-phoneme corrrespondence rules (Venezky, 1970) than for words that are exceptions to those rules (Baron & Strawson, 1976; Edgmon, cited by Gough & Cotsky, 1977; Stanovich & Bauer, 1978; Barron, 1978). Similarly, when grapheme-phoneme regularity is redefined in terms of the consistency of an orthographically specifed neighborhood (Glushko, 1979, 1981), words from phonologically consistent neighborhoods are recognized faster in lexical decision than are words from phonologically inconsistent neighborhoods (Bauer & Stanovich 1980). Here, it is assumed that only when the grapheme-phoneme correspondences are consistent and regular is a phonologically analytic strategy appropriate. If recognition were exclusively dependent on the lexicon, then as long as word frequency were controlled, regular words should not be faster than exception words. The assumption here is that regular words are faster than exception words because there is an advantage to operating a frequency-sensitive word-specific strategy and a phonologically-analytic word-nonspecific strategy together.

Early support for a phonological strategy was derived from the detriment to performance on lexical decision with word homophone letter strings such as weak/week and pseudoword homophone strings such as burd and blud (Rubenstein, Lewis, & Rubenstein, 1971). Later replications (Coltheart, Davelaar, Jonasson, & Besner, 1977) found that the effect of homophony was tied to lexical search in that it only occurred for the lower frequency word in the homophonic pair and that the visual similarity of the pseudoword (but not the real word) to other real words affected reaction time. (Similarity was defined by how many words could be produced by changing any one letter in the pseudoword.) Generally, the detriment due to homophony, as evidence of a phonological strategy, is more robust for pseudowords than for words. As Coltheart et al. (1977) point out, however, the failure to find effects of

homophony for words indicates that possible lexical entries are not both searched in a serial fashion (from high to low frequency) and phonologically specified. Alternatively, failure to demonstrate evidence of a phonological strategy might reflect readers' skill level or the constraints on strategy imposed by the experimental task.

From a developmental perspective, good beginning readers were slower with pseudoword homophones than with control items, while poor readers performed equally with both types of letter strings (Barron, 1978). While poor readers may never employ a phonological analysis, skilled readers can use a phonological recognition strategy, although it is optional and may be suppressed when necessary. With skilled readers, a detriment to performance does occur for the lower frequency homophone word (e.g., altar, beech) when the accompanying pseudowords are not homophones of real words (e.g., slint). If the pseudowords are homophones of real words, however (e.g., brane, brume), then subjects can suppress a phonological stategy (Davelaar, Coltheart, Besner, & Jonasson, 1978; McQuade, 1981).

The effect of homophony, like the influence of phonological consistency in orthographic neighborhoods, is often treated as a post-lexical condition, resulting from a mismatch between a letter string and one (or several) lexical entries (Bauer & Stanovich, 1980). This account assumes an interference due to the inconsistent phonological descriptions provided by different (word-specific) lexical entries. It is not necessary that the knowledge structure of plausible phonological interpretations for multi-letter units be word-specific, however. And, to the extent that these phonological effects occur among pseudowords, they cannot be lexically derived.

In most conceptualizations, the strategies operate simultaneously and interdependently with the assumption that either of the phonological strategies generally acts more slowly than the word-specific strategy. Thus, the latency difference between words and pseudowords is explained: Responding by a visual strategy, an option that is only viable for words, will be faster than responding by a phonological strategy, such as would be necessitated by pseudowords (e.g., Meyer, Schvaneveldt, & Ruddy, 1974; Coltheart et al., 1977; Coltheart et al., 1979). Likewise, phonological effects will be more easily demonstrated with pseudowords than with words.

For words in English, Coltheart et al. (1979) have claimed that the phonological strategies are always optional, but the word-specific visual strategy is sometimes mandatory. From the perspective of task, this word-specific strategy is not necessary for naming, while the phonological strategies may or may not contribute to lexical decision. Henderson (1977) has claimed that the participants in the reading debate have not adequately considered the preservation of morphology in the orthography (but see Taft & Forster, 1975). In support of this, there is a suggestion that within the experimental setting, the number of morphemes in a word affects recognition strategy (Rubin, Becker, & Freeman, 1979). All of the studies on word recognition mentioned above were conducted in English, but it is possible that some of these results reflect peculiarities of English and do not apply to reading in other languages. It therefore becomes essential to test the dominant theory of word recognition and reading in languages that differ from English in the relation between phonology, morphology, and the written form.

Serbo-Croatian: A Phonologically Shallow Orthography

In contrast to the English orthography, which tends to be morphophonemic in its referent (Chomsky, 1970), the writing system of Serbo-Croatian preserves a very close relation to (classical) phonemics and reveals morphological relatedness only when the phonology is similar. In Serbo-Croatian, all similar orthographic patterns will sound alike. Even fully systematic phonological alternations in surface forms are represented in the orthography so that visual or orthographic similarity of morphologically related forms may be obscured; for example, nominative singular RUK+A, dative singular RUC+I; nominative singular SNAH+A, dative singular SNAS+I. (Note: Inflection is the major grammatical device of Serbo-Croatian. The preceding are Roman transcriptions of the English words, ARM and DAUGHTER-IN-LAW, respectively.) In addition, as a result of the tendency toward open syllables, the possible patterning of consonants and vowels is much more restricted in Serbo-Croatian than in English. Not only do the orthotactic (Taft, 1979) rules fully mimic the phonotactic rules, but the possibility for ambiguous syllable boundaries due to sequences of consonants is greatly reduced.

The depth of an alphabetic orthography is reflected by the extent to which the spoken form is specified by the orthographic form: That is, by the complexity of the derivational rules that relate the orthographic transcription to some (abstract) description appropriate for speaking. A deep orthography with a complex relation to the spoken form may induce a word-specific strategy that avoids the derivations. In English, the complex relation between written and spoken form is increased because, historically, the written form and the speech form have not evolved in the same way. Therefore, the graphemic transcription often does not correspond exactly to the phonology and this could influence recognition strategy.

In comparison with the derivational rules for English, Serbo-Croatian has maintained a close correspondence between the written and spoken forms. This is the outcome of deliberate alphabet reforms introduced by Karadžić and Gaj in the last century that reconstructed the Roman and Cyrillic alphabets in which the Serbo-Croatian language is written according to the simple rule: "Write as you speak and speak as it is written." As a result, the Roman and the Cyrillic orthographies transcribe the sounds of the Serbo-Croatian language in a direct and consistent manner, and there are no (nontrivial) derivational rules. In summary, the orthography is shallow and there are no exception words in Serbo-Croatian. Consequently, a word-specific strategy would never be required.

Since the Roman and Cyrillic alphabets transcribe the same language, their graphemes must map onto the same set of phonemes. These two sets of graphemes are, with certain exceptions, mutually exclusive (see Table 1). Most of the Roman and Cyrillic letters are unique to their respective alphabets. There are, however, a number of letters that the two alphabets have in common. The phonemic interpretation of some of these shared letters is the same whether they are read as Cyrillic or as Roman graphemes; these are referred to as <u>common</u> letters. Other members of the shared letters have two phonemic interpretations, one in the Roman reading and one in the Cyrillic reading; these are referred to as <u>ambiguous</u> letters (see Figure 1).

TABLE 1

SERBO-CROATIAN				
ROMAN		CYRILLIC		
PRINTED		PRINTED		LETTER NAME IN I.P.A.
UPPER CASE	LOWER CASE	UPPER CASE	LOWER CASE	
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	nə
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

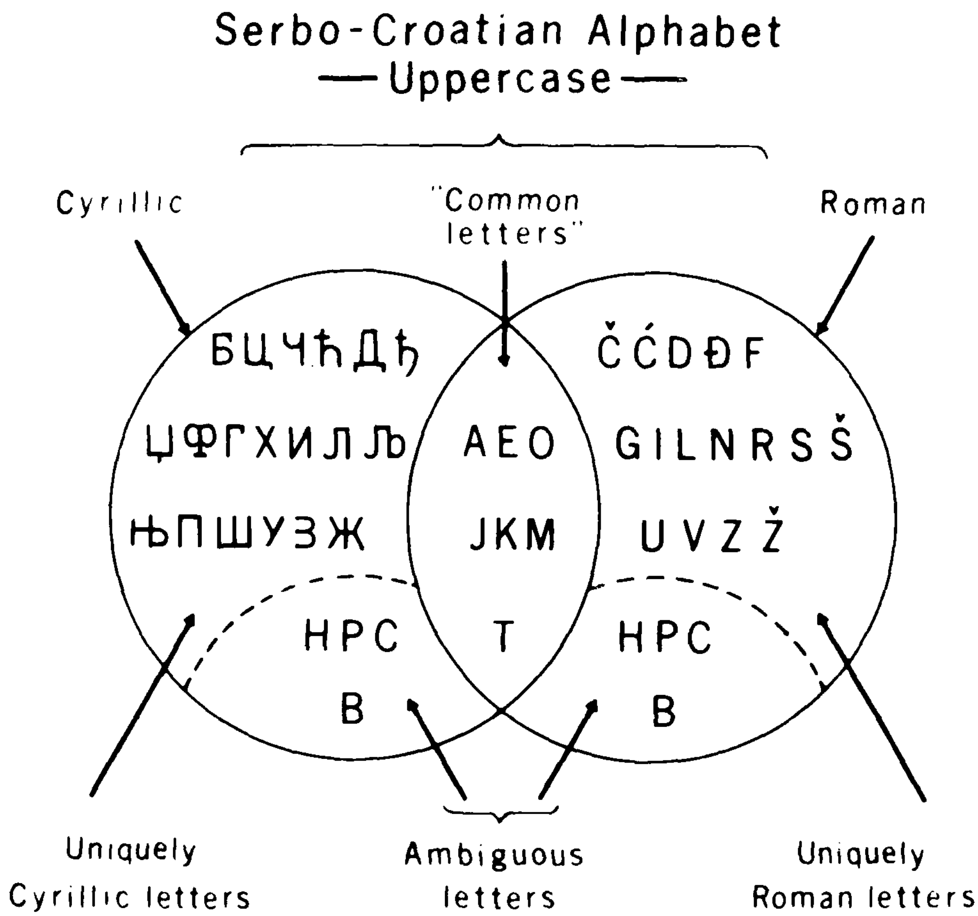


Figure 1. Letters of the Roman and Cyrillic alphabets.

Given the nature of and the relation between the two Serbo-Croatian alphabets, it is possible to construct a variety of types of letter strings. A letter string of uniquely Roman and common letters or of uniquely Cyrillic and common letters would be read in only one way and could be either a word or nonsense. A letter string composed of the common and ambiguous letters could be pronounced in one way if read as Roman and pronounced in a distinctively different way if read as Cyrillic; moreover, it could be a word in one alphabet and nonsense in the other or it could represent two different words, one in one alphabet and one in the other, or finally, it could be nonsense in both alphabets (see Table 2).

Whatever their category, the individual letters of the two alphabets have phonemic interpretations that are virtually invariant over letter contexts. Moreover, all the individual letters in a string of letters, be it a word or nonsense, are pronounced--there are no letters made silent by context.[1] Finally, but not least in importance, a large portion of the population uses both alphabets competently. This is due, in part, to an education requirement that both alphabets be taught within the first two grades. The Roman alphabet is taught first in the western part of Yugoslavia and the Cyrillic alphabet is taught first in the eastern part of Yugoslavia.

In sum, the Serbo-Croatian orthography relative to the English orthography permits less variability in its orthotactic patterning relative to phonotactic patterns, but more variability in the written form of some base morphemes. It is less concerned with preserving morphological relatedness and closely relates to the spoken language. The depth of an orthography reflects the extent to which the phonetic rendition is specified by the orthographic form: Serbo-Croatian is characterized as a shallow orthography.

Word Recognition in Serbo-Croatian

The complex relation between letter and sound in English reflects its phonologically deep orthography and the opaqueness of this relation is offered as a reason why phonological involvement in the fluent reading of English is not efficient (Goodman, 1976; Kolers, 1970; Smith, 1971). This reasoning would not preclude a phonological strategy in the fluent reading of Serbo-Croatian, however. Due to the systematic relation of graphemes and phonemes, in principle, a reader of Serbo-Croatian could arrive at a phonological description of a word correctly without ever relying on knowledge about the specific word.[2] Differences among orthographies in structure and in phonological depth may influence reading strategies, in which case a model of word recognition delineated for English may prove inadequate when applied to Serbo-Croatian.

The shallow character of the Serbo-Croatian orthography rationalizes a phonological priority relative to a word-specific priority in reading and word recognition and there is empirical support for this claim. In Serbo-Croatian, an effect detrimental to performance on lexical decision with phonologically bivalent grapheme strings was demonstrated for words (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) and later, for both words and pseudowords (Lukatela, Popadić, Ognjenović, & Turvey, 1980). In the earlier experiment, both the design of the experiment and the instructions to the subjects were selected to restrict the task to the Roman alphabet, but subjects were unable

Table 2

Types of Letter Strings and Their Lexical Status

Composition of Letter String	Phonemic Interpretation	Meaning
AMBIGUOUS and COMMON		
CABAHA*	Cyrillic /savana/	savanna
	Roman /tsabaxa/	nonsense
KOBAC	Cyrillic /kovas/	nonsense
	Roman /kobats/	hawk
KACA	Cyrillic /kasa/	safe
	Roman /katsa/	pot
HEPETAC*	Cyrillic /neretas/	nonsense
	Roman /xepetats/	nonsense
COMMON		
JAJE	Cyrillic /jaje/	egg
	Roman /jaje/	egg
TAKA	Cyrillic /taka/	nonsense
	Roman /taka/	nonsense
UNIQUE and COMMON		
SAVANA*	Cyrillic impossible	
	Roman /savana/	savanna
NERETAS*	Cyrillic impossible	nonsense
	Roman /neretas/	nonsense
КОБАЦ	Cyrillic /kobats/	hawk
	Roman impossible	
ПУДАЛ	Cyrillic /pudal/	nonsense
	Roman impossible	

(*indicates those letter string types included in the present experiment)

to suppress a Cyrillic alphabet reading when the letter string permitted one. In the later experiment (Lukatela et al., 1980), no alphabet restriction was imposed. This detriment could be interpreted as either a (visual) alphabet or a phonology-induced ambiguity. Because the phonological bivalence of those ambiguous graphemes should exert no influence on visual matching and because those words composed of shared characters with a common phonemic value in both alphabets were no slower than pure Roman strings, it was concluded that for the phonologically shallow orthography of Serbo-Croatian, lexical decision always proceeds with reference to phonology. Not only was the effect replicated for pseudowords (Lukatela et al., 1980), but the influence of phonological bivalence on words occurred both when the alternate reading produced a word or a pseudoword (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978). Therefore, this effect is not easily characterized in terms of the differing lexical status of the alternate reading. In that experiment, however (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978), subjects responded positively only to those letter strings that were words in Roman. Therefore, words in Cyrillic, as well as all pseudowords, required a negative response. A better test of the influence of lexical status of the alternate readings is currently underway (Feldman, Note 1).

The present work continues to investigate whether the phonological coding strategy for word recognition is optional in the phonologically shallow orthography of Serbo-Croatian. In the original bivalent phonology experiments (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978; Lukatela et al., 1980), different words occurred in the phonologically unique and phonologically bivalent conditions. Therefore, the effect of phonological bivalence was assessed between words. Although word frequency range was balanced across conditions, the effect of a unique or a bivalent phonology was measured on different letter strings. In sum, evidence for a phonological recognition strategy for lexical decision on words has been demonstrated for Serbo-Croatian by comparing between different word types. In the present experiment, the internal orthographic structure of the letter string was constructed in such a way that the punitive effect of phonological coding could be assessed within (two forms of the same) words for both the naming and lexical decision tasks.

As discussed above, there are two possible strategies or codes by which access to the lexicon or the process of word recognition can occur. If, as sometimes implied for English, there is only one phonological code and if this phonological description is lexically derived such that word identification must rely on some familiar visual form or an unanalyzed pattern, then word recognition should be independent of phonological factors and be closely tied to a holistic orthographic form. In this case, effects of phonological variables should not impair (or facilitate) word performance on linguistic tasks such as lexical decision. This word-specific strategy is differentially effective according to lexical status. For words, either the word-specific strategy or its secondary phonological strategy could operate in principle. For pseudowords, however, a phonological strategy is the only possibility, as the pseudowords are not familiar and have not been encoded previously. As a result, a word-specific strategy would predict that variables that introduce phonological complexity should have a greater effect on pseudowords than on words (see Coltheart, 1978).

To the extent that phonological strategies are sensitive to the components of orthographic structure and to the position of phoneme clusters within the letter string (Frederiksen & Kroll, 1976), the impairment due to phonological bivalence should vary as a function of the number and distribution of ambiguous characters within the letter string. In a recent experiment (Feldman, Kostić, Lukatela, & Turvey, 1981, this volume), overall effect of a phonologically bivalent sequence of letters could be alleviated if a unique letter appeared in the final position. In addition, and more important to the present investigation, in a fully ambiguous string the magnitude of the impairment depended on the number of ambiguous characters. In that experiment, all comparisons were <u>within</u> words in that they were made on the difference between the ambiguous and unique readings of the same word (or morpheme-based unit). Therefore, there was no contamination due to word frequency, word length, or richness of meaning. While there is evidence that skilled readers in Serbo-Croatian exploit syllable units (Katz & Feldman, 1981), in the experiment reported by Feldman et al. (1981), number of ambiguous letters was confounded with number of ambiguous syllables and all letter strings had two syllables.

In the present experiment, the within-word effect of phonological bivalence on naming was investigated and the effect or lexical decision was replicated. If phonological bivalence impairs performance for both words and pseudowords, then these results would suggest that a phonological strategy is mandatory regardless of the lexical status of the letter string. If the impairment due to phonological bivalence is greater for words than for pseudowords, then the notion of a phonological strategy employed only as the complement of a word-specific strategy is invalidated. If the effect obtains for naming as well as for lexical decision such that a correlation is obtained between latencies in the two tasks, then a common knowledge structure must participate in both tasks. And, if the effect of phonological bivalence varies with the number or position of the ambiguous letters within the string, then a phonographically analytic phonological strategy must be operative.

METHODS

Subjects

Sixty-two first year students of psychology at the University of Belgrade participated in this study in partial fulfillment of course requirements. Twenty-eight subjects performed lexical decision judgments and thirty-four subjects performed a naming task. Subjects were eliminated from the study if their error rate exceeded 10%. This occurred with six subjects in the naming portion. In all, there were 56 subjects, 28 in each task, whose data were included in the statistical analysis.

Stimuli

Each subject viewed 246 slides, which included 30 practice trials. Half of the letter strings were words and half were pseudowords that were actually derived from other real words by changing two or three letters in the latter portion of the letter string. Half of the items contained two syllables (with five or six letters) and half contained three syllables (with six or seven

letters). All words were nouns in the mid-frequency range as judged by consensus among several native speakers. Each subject saw three types of words and pseudowords defined by the manner in which they were presented across subject groups. CONTROL items were printed in Roman for both groups of subjects, e.g., MUZIKA. PURE items were printed in Cyrillic for half the subjects (Group One) and in Roman for the other half (Group Two). These PURE letter strings contained characters that are unique to an alphabet (either Cyrillic or Roman), in both their Roman and their Cyrillic transcriptions. The third type were AMBIGUOUS items, chosen such that they contain only common and ambiguous characters in the Cyrillic rendition. In contrast, in the Roman version, these letter strings contain characters that are unique to the Roman alphabet. As a result, the Cyrillic form permits two different readings while the Roman form specifies a unique reading. Within the ambiguous letter strings, number and position of ambiguous characters were systematically varied. For the three syllable items, two or three ambiguous characters were distributed over two or three syllables. For the two syllable items, one or two ambiguous characters were distributed over one or two syllables (see Table 3).

Procedure

Twenty-eight subjects performed a lexical decision task. As each word appeared, they had to tap a key with both hands to indicate "yes" (further key) or "no" (closer key), in deciding whether or not each stimulus was a word. The other twenty-eight subjects performed the naming task. That is, they had to read each word aloud as rapidly as possible. All stimuli were typed on Prima U Film and the Cyrillic and Roman typeface were closely matched for size and form. (Common characters were identical in the two typefaces.) In contrast to the lexical decision task, responses in the naming task were timed with a voice-operated relay that began counting with the onset of the visual display.

In the instructions for lexical decision, subjects were informed that words would appear both in Roman and in Cyrillic. During the experimental session, subjects were advised of their mistakes. In the naming task, subjects were given the same description of the stimulus set as in the lexical decision task. They were instructed to pronounce each string as a word if it could be read as such. For all subjects, stimuli were presented for 750 msec in one channel of a Scientific Prototype model GB Tachistiscope. A blank field immediately preceded and followed the display interval. The interval between experimental trials was about 2000 msec and reaction times were measured from the onset of the stimulus display. A brief pause was introduced halfway through the experimental session.

Each group of subjects saw eighteen Cyrillic words and eighteen Cyrillic pseudowords intermixed with ninety Roman words and ninety Roman pseudowords. For both lexical decision and naming, Group Two subjects saw eighteen AMBIGUOUS Cyrillic words, e.g., CABAHA (/savana/) (which could also be read as a pseudoword in Roman /tsabaxa/) and eighteen PURE words in Roman, e.g., FABRIKA (/fabrika/), as well as eighteen AMBIGUOUS Cyrillic pseudowords, e.g., HEPETAC (/neretas/ or /xepetats/) and eighteen PURE pseudowords in Roman, e.g., EDOGOM (/edogom/). In addition, they saw a CONTROL set of seventy-two words and seventy-two pseudowords written in Roman. Group One subjects saw

Table 3

Distribution of Ambiguous Letters and Phonemic Interpretation for AMBIGUOUS Cyrillic Letter Strings

Three Syllable Letter Strings	Phonemic Interpretation	Meaning	Number of Ambiguous Letters	Number of Ambiguous Syllables
CABAHA	Cyrillic /savana/	savanna	3	3
	Roman /tsabaxa/	nonsense		
KAPABAH	Cyrillic /karavan/	caravan	3	2
	Roman /kapabax/	nonsense		
OCTABKA	Cyrillic /ostavka/	resignation	2	2
	Roman /otstabka/	nonsense		
Two Syllable Letter Strings				
OPMAH	Cyrillic /orman/	cabinet	2	2
	Roman /opmax/	nonsense		
CAHTA	Cyrillic /santa/	iceberg	2	1
	Roman /tsaxta/	nonsense		
KOTBA	Cyrillic /kotva/	anchor	1	1
	Roman /kotba/	nonsense		

the same AMBIGUOUS words, now written in Roman, e.g., SAVANA, where they are no longer ambiguous, and the PURE words written in Cyrillic, e.g., ФАБРИКА, as well as eighteen AMBIGUOUS pseudowords written in Roman and eighteen PURE pseudowords written in Cyrillic. Group One, like Group Two, saw the CONTROL words and pseudowords written in Roman, e.g., MUZIKA.

In summary, for both the lexical decision and naming tasks there were two groups of subjects. The PURE Cyrillic words (18) and pseudowords (18) for Group One were presented to Group Two in Roman and the unique Roman version of the AMBIGUOUS words (18) and pseudowords (18) from Group One were presented to Group Two in their AMBIGUOUS Cyrillic form. In addition, both groups saw the same set (72 each) of Roman words and of pseudowords. As a result, the ratio of Cyrillic words to Roman words was one to five for both groups of subjects. All comparisons between groups were therefore performed on the same set of words where the alphabet changes (for the PURE and for the AMBIGUOUS word sets) across subject groups.

As noted above, if Group One saw a particular word type in its Roman version, then Group Two saw that same word type in an AMBIGUOUS Cyrillic version. Conversely, the PURE Cyrillic word type from Group One appeared in Roman for Group Two. The two types of Cyrillic words differ in one important respect: The Cyrillic words for Group Two, i.e., AMBIGUOUS words, are also readable in Roman. This is not true for the other type, the PURE words, which were presented to Group One. Phonological bivalence is restricted to Group Two's Cyrillic words and pseudowords.

RESULTS

Lexical Decision

An analysis of variance for lexical decision, with minimum and maximum latencies set at 250 msec and 2500 msec, revealed highly significant effects for lexicality (word/pseudoword), min $F'(1,21) = 21.15$, $p < .001$; for group (one/two), min $F'(1,15) = 20.28$, $p < .001$; for word type (ambiguous/pure/control), min $F'(2,16) = 22.35$, $p < .001$; and for length in syllables (two/three), min $F'(1,11) = 6.22$, $p < .05$. In addition, the type x group interaction was significant with min $F'(2,16) = 20.73$, $p < .001$. The lexicality x type x group interaction was also significant with min $F'(2,20) = 6.66$, $p < .01$.

Mean number of errors per subject for lexical decision was 4 for Group One and 12 for Group Two. Considering only the ambiguous type items, mean errors were 2 for Group One and 8 for Group Two (see Table 4). For all items for both groups, there was no evidence of a speed-accuracy trade-off. In fact, reaction time and errors were positively correlated; for Group One, $r = .33$, for Group Two, $r = .50$. These correlations were significantly different, $z = 2.09$, $p < .05$, but the difference is most likely due to the restricted range of scores for Group One. In order to assess the possibility that subjects altered their strategy as they proceeded through the task, the correlation of the difference between the unique Roman and the ambiguous Cyrillic latency for each word (and pseudoword) and position of the item in the list was computed for each item. (A large number indicates a position

Table 4

Summary of
Data for Lexical Decision on AMBIGUOUS Cyrillic/Unique
Roman Letter Strings

LENGTH IN SYLLABLES LEXICALITY	TWO WORD	TWO PSEUDOWORD	THREE WORD	THREE PSEUDOWORD
ROMAN	ORMAN	VAMAS	SAVANA	NERETAS
MEAN	632	717	677	769
STANDARD DEVIATION	86	76	89	62
ERRORS	.4	.3	.7	.6
CYRILLIC	OPMAH	BAMAC	CABAHA	HEPETAC
MEAN	945	925	984	993
STANDARD DEVIATION	106	144	123	139
ERRORS	3.3	.5	3.9	.4

late in the list.) For lexical decision, the correlation was not significant ($r = .19$). This result suggests that reliance on a phonological strategy did not diminish during the experimental session. Similarly, in order to assess the possibility that reliance on a phonological strategy varied with word frequency, the reaction time to the unique Roman version of each word was used as an estimate of word frequency, and the correlation between unique Roman latency and the difference between the unique Roman and ambiguous Cyrillic form of each word was computed. In lexical decision, the correlation was not significant ($r = -.17$). Therefore, reliance on a phonological strategy did not vary as a function of word frequency.

Protected t-tests between mean reaction times for lexical decision (with the estimate of variance computed from the subject's analysis of variance) showed that the significant interactions of type x group and type x group x lexicality could be attributed to a significant difference between AMBIGUOUS Cyrillic/unique Roman form of words, (CABAHA/SAVANA), $t(13) = 8.89$, $p < .001$ (see Figure 2). Groups did not differ significantly on uniquely Cyrillic or Roman PURE words, (ФАБРИКА/FABRIKA), $t(13) = 1.09$. Therefore, there is no general tendency for Roman items to be recognized more quickly than the Cyrillic version of those same items. The between-group difference on CONTROL words (MUZIKA) only approached significance, $t(13) = 1.96$, $p < .10$. Nevertheless, the magnitude of the AMBIGUOUS and CONTROL word difference across groups varied significantly, $t(13) = 7.91$, $p < .001$. The unique Roman and the ambiguous Cyrillic forms of the AMBIGUOUS type words differed more than the (consistently) Roman forms of CONTROL words. Pseudowords demonstrated a smaller effect of ambiguity than did words, $t(13) = 3.6$, $p < .001$. For Group One the difference between (unique) word types was not significant, while for Group Two, the difference between word types was significant. Group Two was always slower than Group One; however, the magnitude of the difference between groups varied over word types. Finally, ambiguous type pseudowords differed more in their Roman and Cyrillic forms than did PURE type pseudowords, $t(13) = 3.74$, $p < .01$.

In order to ascertain the effect of ambiguous characters, another analysis of variance was performed including only the ambiguous Cyrillic and unique Roman forms of the AMBIGUOUS type words and pseudowords. Because of the special constraints on selecting these words, no Clark analysis (1973) was performed. Instead, the results of an analysis of variance using subject variability as the error term(s) are reported.

In this analysis, letter strings were classified according to the number and distribution of ambiguous characters within the letter string. As in the more complete lexical decision analysis discussed above, there were significant main effects of group, $F(1,26) = 99.44$, $MS_e = 159087$, $p < .001$, and length of word in syllables, $F(1,26) = 9.62$, $MS_e = 11117$, $p < .01$. In contrast to previous analyses, however, lexicality only approached significance, $F(1,26) = 2.48$, $MS_e = 57878$, $p < .20$. Importantly, the distribution x group interaction was significant, $F(2,52) = 4.88$, $MS_e = 8398$, $p < .05$, as was the distribution x group x lexicality interaction, $F(2,52) = 10.55$, $MS_e = 218937$, $p < .01$.

Protected t-tests on the within-word difference between means for the unique Roman and ambiguous Cyrillic transcription of the same words (pooled

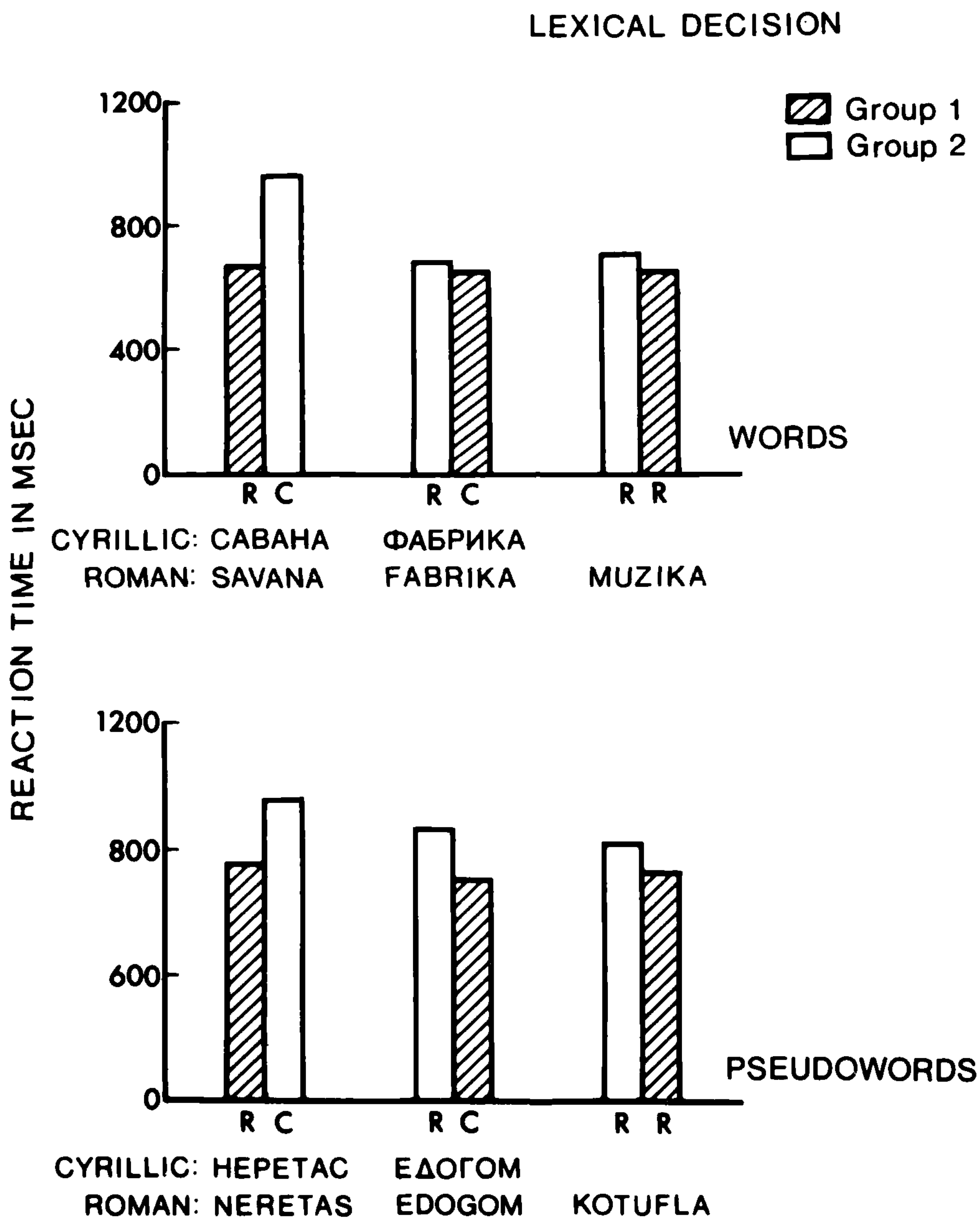


Figure 2. Mean reaction time for lexical decision on AMBIGUOUS (CABAHA), PURE (FABRIKA) and CONTROL (MUZIKA) words and pseudowords written in Roman and in Cyrillic.

over two- and three-syllable words) revealed that when number of ambiguous syllables was controlled, number of ambiguous characters increased latencies significantly, $t(13) = 3.65$, $p < .01$. And when number of ambiguous characters was controlled, clustering two ambiguous characters within one syllable was more difficult than having the two ambiguous letters distributed through different syllables, $t(13) = 2.62$, $p < .05$ (see Table 5). For pseudowords, none of the contrasts among various distributions of ambiguous letters was significant.

An analysis of variance conducted on the errors in judging the lexical status of unique Roman and ambiguous Cyrillic forms of the AMBIGUOUS word type provided the same basic results as did the reaction time analysis: Main effects of lexicality and group were significant, as was their interaction; $F(1,26) = 38.20$, $MS_e = 65.93$, $p < .001$; $F(1,26) = 39.08$, $MS_e = 56.32$, $p < .001$, and $F(1,26) = 31.85$, $MS_e = 65.93$, $p < .001$, respectively. Here, length of the word in syllables was not significant, $F(1,26) = 2.16$, $MS_e = 27.0$, $p > .20$. And, in the error analysis, the distribution of ambiguous characters was not significant, $F(2,52) = 1.0$, $MS_e = 12.78$ and did not interact with group, although it did interact with other variables: distribution x syllable, $F(2,52) = 3.25$, $MS_e = 25.41$, $p < .05$; distribution x syllable x lexicality, $F(2,52) = 17.74$, $MS_e = 34.48$, $p < .001$; distribution x lexicality x group, $F(2,52) = 3.82$, $MS_e = 21.29$, $p < .05$.

Naming

In the analysis of variance performed on the naming data, with the same criteria for minimum and maximum latencies, a very similar pattern emerged. There were highly significant results for lexicality: min $F'(1,16) = 50.49$, $p < .001$; for group, min $F'(1,15) = 20.76$, $p < .001$; for word type, min $F'(2,12) = 45.55$, $p < .001$, and for length in syllables, min $F'(1,10) = 29.04$, $p < .001$. As above, the type x group interaction was significant, min $F'(2,20) = 90.96$, $p < .001$, but in contrast to the lexical decision results, the lexicality x group interaction was also significant, min $F'(1,20) = 7.81$, $p < .05$, and the lexicality x type x group interaction was not: min $F'(2,11) = .08$.

In naming, mean number of errors per subject was 11 in Group One and 15 in Group Two. For the ambiguous word type alone, mean number of errors was 2 in Group One and 9 in Group Two (see Table 6). Once again, there was no evidence of a speech-accuracy trade-off. The correlation between reaction time and errors for both word and pseudoword items was $r = .61$ for each group. When the difference between the unique Roman and the ambiguous Cyrillic form of each word was correlated with the position of the item within the list, the correlation ($r = .19$) was not significant. These data suggest that reliance on a (detrimental) phonological strategy did not diminish during the experimental session. As in the lexical decision task, in order to examine whether reliance on a phonological strategy varies with word frequency, reaction time to the unique Roman form of each word was treated as an index of word frequency, and a correlation between the unique Roman form and the difference between the unique Roman and the ambiguous Cyrillic form of each word was computed. In the naming task, in contradistinction to the lexical decision task, this correlation was significant and positive, $r = .54$, $p < .05$. It suggests that the detriment due to phonological bivalence decreases with frequency.

Table 5

Mean Reaction Time by Distribution of Ambiguous Characters for Lexical Decision on AMBIGUOUS Cyrillic Words (and Their Roman Controls)

Three Syllable Letter Strings	Number of Ambiguous Letters	Number of Ambiguous Syllables	Cyrillic Reaction Time	Roman Reaction Time	Difference Between Cyrillic and Roman
CABAHA	3	3	981	676	305
KAPABAH	3	2	1038	646	392
OCTABKA	2	2	934	709	245
Two Syllable Letter Strings					
OPMAH	2	2	927	645	273
CAHTA	2	1	1027	650	377
KOTBA	1	1	880	625	255

Table 6

Summary of Data to Name
AMBIGUOUS Cyrillic/Unique Roman Letter Strings

LENGTH IN SYLLABLES / LEXICALITY	TWO WORD	TWO PSEUDOWORD	THREE WORD	THREE PSEUDOWORD
ROMAN	ORMAN	VAMAS	SAVANA	NERETAS
MEAN	621	668	686	724
STANDARD DEVIATION	41	83	66	59
ERRORS	.1	.1	.1	1.4
CYRILLIC	OPMAH	BAMAC	CABAHA	HEPETAC
MEAN	1009	1194	1132	1258
STANDARD DEVIATION	207	248	166	204
ERRORS	2.2	2.1	1.8	2.9

Protected t-tests on mean naming latencies (with the mean square error term derived from the subjects' analysis of variance) confirmed that latencies to name ambiguous words and pseudowords were prolonged. While there was no significant difference between groups on Roman CONTROL words, $t(13) = 1.51$, groups did differ on AMBIGUOUS Cyrillic/unique Roman words, $t(13) = 14.95$, $p < .001$. And the difference between groups was greater for the AMBIGUOUS type items than for the PURE type items, $t(13) = 13.45$, $p < .001$. In contrast to the lexical decision data, the difference between naming ambiguous Cyrillic and unique Roman appeared greater for pseudowords than for words, $t(13) = 4.01$, $p < .01$. This result is difficult to evaluate (and a protected t-test is not strictly legal) since the lexicality x group type interaction was not significant. Because there was no "correct" reading of an ambiguous pseudoword, both Cyrillic and Roman readings were included in the analysis, and, in fact, this condition had a larger variance than its unique Roman counterpart. (Standard deviations for ambiguous Cyrillic pseudowords of two and three syllable length were 248 and 204, respectively, while their Roman equivalents were 83 and 59.) Finally, Group Two was slower on Ambiguous Cyrillic than on Pure Roman strings, $t(13) = 15.08$, $p < .001$. For Group One, there was no evidence of an alphabet bias as Pure Cyrillic and Unique Roman string were equal, $t(13) = .72$ (see Figure 3).

In order to evaluate the effect of the distribution of ambiguous characters on naming, an analysis of variance including only the ambiguous Cyrillic and unique Roman naming latencies of the AMBIGUOUS type words and pseudowords was also performed. As in the analogous lexical decision analysis, a Clark analysis (1973) was not appropriate due to the severe selection constraints on words. Therefore, the results reported below are based on an analysis of variance of naming latencies using subject variability as the error term(s).

In agreement with the larger naming analysis discussed above, there were significant main effects of group, $F(1,26) = 89.54$, $MS_e = 210297$, $p < .001$, length in syllables, $F(1,26) = 34.41$, $MS_e = 14409$, $p < .001$, and lexicality, $F(1,26) = 68.32$, $MS_e = 12020$, $p < .001$, and a significant group x lexicality interaction, $F(1,26) = 21.86$, $MS_e = 12020$, $p < .001$. When letter strings were classified according to the number and distribution of ambiguous characters (distribution), distribution was significant, $F(2,52) = 5.31$, $MS_e = 11313$, $p < .01$, as were the interactions of distribution x syllable, $F(2,52) = 12.07$, $MS_e = 10582$, $p < .001$; distribution x lexicality, $F(2,52) = 4.80$, $MS_e = 14941$, $p < .05$ and, most important, distribution x group, $F(2,52) = 3.48$, $MS_e = 11313$, $p < .05$. The second order interactions of distribution x syllable x group and distribution x syllable x lexicality were also significant: $F(2,52) = 7.55$, $MS_e = 10582$, $p < .01$ and $F(2,52) = 14.40$, $MS_e = 10626$, $p < .001$, respectively. Finally, the fourth order interaction of distribution x lexicality x group x syllable was also significant, $F(2,52) = 19.78$, $MS_e = 10626$, $p < .001$.

Protected t-tests on the naming data resembled the results for lexical decision. For words, when number of ambiguous characters was controlled, two ambiguous characters within one syllable were slower than one ambiguous character within a syllable, $t(13) = 2.54$, $p < .05$ for three-syllable words, and $t(13) = 2.82$, $p < .05$ for two-syllable words (see Table 7). There were no other significant results for words and, probably due to their large variances, no significant results for pseudowords.

AD-A105 090 HASKINS LABS INC NEW HAVEN CT F/G 17/2
STATUS REPORT ON SPEECH RESEARCH, 1 APRIL-30 JUNE 1981.(U)
1981 PHS-N01-HD-1-2420
UNCLASSIFIED SR-66(1981) NL

3 OF 4
AD A
105090

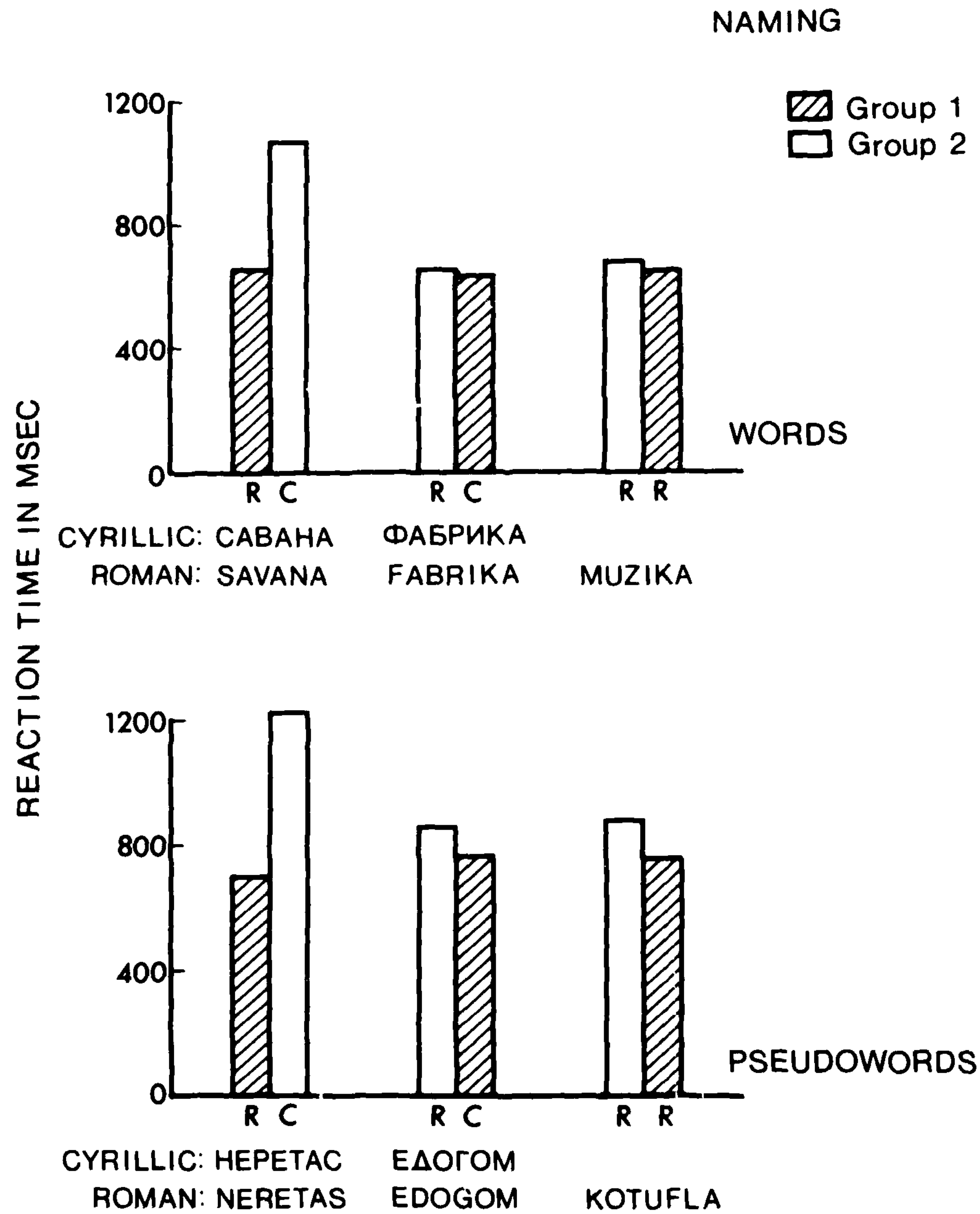


Figure 3. Mean reaction time to name AMBIGUOUS (CABAHA), PURE (FABRIKA) and CONTROL (MUZIKA) words and pseudowords written in Roman and in Cyrillic.

Table 7

Mean Reaction Time by Distribution of Ambiguous Characters to Name AMBIGUOUS Cyrillic Words (and Their Roman Controls)

Three Syllable Letter Strings	Number of Ambiguous Letters	Number of Ambiguous Syllables	Cyrillic Reaction Time	Roman Reaction Time	Difference Between Cyrillic and Roman
CABAHA	3	3	1049	661	388
KAPABAH	3	2	1047	609	438
OCTABKA	2	2	933	594	339
Two Syllable Letter Strings					
OPMAH	2	2	1125	703	422
CAHTA	2	1	1201	687	514
KOTBA	1	1	1071	667	404

An analysis of variance conducted on errors in naming the unique Roman and ambiguous Cyrillic forms of the AMBIGUOUS word type was generally consistent with the reaction time results. Significant main effects of group, $F(1,26) = 23.83$, $MS_e = 129.9$, $p < .001$, and lexicality, $F(1,26) = 4.82$, $MS_e = 63.23$, $p < .05$ occurred, although length in syllables was not significant, $F(1,26) = 3.04$. In contrast to the results for lexical decision, the distribution of ambiguous characters in the error analysis was significant, $F(2,52) = 31.18$, $MS_e = 20.63$, $p > 001$, and distribution interacted with group, $F(2,52) = 7.06$, $MS_e = 20.63$, $p < .01$. The interaction of distribution x syllable and distribution x syllable x lexicality were also significant, $F(2,52) = 9.79$, $MS_e = 26.49$, $p < .001$ and $F(2,52)5.16$, $MS_e = 27.77$, $p < .01$, respectively, as was the interaction of distribution x syllable x lexicality x group, $F(2,52) = 12.96$, $MS_e = 27.77$, $p < .001$.

The correlation between means for individual letter strings in the naming and lexical decision tasks was computed for the two groups of subjects who saw the ambiguous Cyrillic letter strings (Group Two's) and separately for the two groups of subjects who saw the unique Roman version of the same items (Group One's). Each correlation was computed on all items, both words and pseudowords, as well as on words alone. While the correlations were slightly higher for words alone than for words and pseudowords combined, these differences were not significant, $z = 1.19$, $p > .25$ for Cyrillic and $z = .69$, $p > .25$ for Roman. Subsequently, all correlations computed between lexical decision and naming included both words and pseudowords, although AMBIGUOUS and CONTROL type items were treated in separate correlations. The correlation between lexical decision and naming was $r = .34$ for unique Roman version of the AMBIGUOUS items (Group One) and $r = .48$ for the AMBIGUOUS Cyrillic version (Group Two). For the Control items, the correlations were $r = .56$ (Group One) and $r = .73$ (Group Two). The difference between the correlations for AMBIGUOUS and CONTROL type items was not significant whether both types of items appeared in Roman, which permitted a unique reading (as for Group One), $z = 1.14$, $p > .25$, or whether the AMBIGUOUS type appeared in its ambiguous form while the CONTROL items uniquely specified a Roman reading (as for Group Two), $z = 1.64$, $p > .10$. These results suggest that the relation between lexical decision and naming did not vary significantly with word type, phonological bivalence, or lexicality.

DISCUSSION

The results of the present experiment demonstrated that phonologically bivalent letter strings retard word recognition relative to the unequivocal form of the same letter string. Similar phonological effects occurred both in naming and in lexical decision, implying a common phonological influence in both tasks. This interpretation was supported by the high correlation between tasks that obtains for both words and pseudowords, and, given the nature of the Serbo-Croatian orthography, it implies a strategy that is not specific to real words alone. Given the nature of the Serbo-Croatian orthography, however, phonological bivalence and visual (alphabetic) bivalence are usually confounded. Before concluding that this effect of phonological bivalence is definitive evidence of a phonological strategy in word recognition, an interpretation in terms of a lexically-based visual search must be invalidated. Most obviously, this detriment occurred for pseudowords as well as words

and it is usually assumed that only words are in the lexicon. To anticipate, allowing that pseudowords as well as words comprise the lexicon and then introducing several further modifications to the lexicon will justify most of the data, but not all: By definition, no account based on a visual search of the lexicon can be sensitive to phonographic analysis of component orthographic structure as the significant effect of the distribution of ambiguous characters would require. It will be concluded that word recognition in Serbo-Croatian is necessarily phonological.

In general, both the lexical decision and the naming paradigms revealed an effect of phonological bivalence that could not be accounted for in terms of any overall difference between subject groups or alphabets. Protected t-tests confirmed that Group One demonstrated no difference between word types and no systematic preference for letter strings in either a Roman or a Cyrillic form. In contrast, Group Two, which was always slower than Group One, was especially impaired on the Ambiguous Cyrillic forms. To the extent that experience with a word in printed text occurs equally often with its Roman and with its Cyrillic form, there should be no difference in latency as alphabet varies. To the extent that the experimental condition can introduce an alphabet bias, this bias should have been similar for both groups and insensitive to word type: The ratio of Cyrillic to Roman items was constant across subjects, all subjects had learned Cyrillic first, and subjects were randomly assigned to experimental groups. And, as predicted, for those subjects who saw PURE (FABRIKA) words in Cyrillic and CONTROL (MUZIKA) and AMBIGUOUS (SAVANA) words in Roman (Group One), there was no difference between alphabets or word types. In assessing any general difference among word types, each contrast entailed a <u>within-words</u> comparison between the Roman and the Cyrillic renditions of the <u>same</u> word displayed to different groups of subjects. Therefore, orthographic and semantic factors as well as word frequency were fully controlled. The effect of phonological bivalence was the difference between the unique Roman and the ambiguous Cyrillic rendition of the same letter string, once any overall difference between groups had been considered. This within-word effect of bivalence was evident in the significant group x type and in the group x type x lexicality interactions.

In summary, the results of the present experiment show that the possibility of two phonological interpretations of a visually presented letter string affected performance on lexical decision and on naming in a way that one phonological interpretation did not. Whether this effect is actually less robust for pseudowords than for words or less reliable in naming than in lexical decision should not confuse the overall conclusion. Latency differences on the order of 300 msec computed on two forms of the <u>same</u> letter string, one phonologically equivocal and one phonologically unequivocal, provide strong evidence of a mandatory phonological strategy in visual word recognition. Before concluding that phonological bivalence need be interpreted as evidence of a phonographically analytic phonological strategy in word recognition, two versions of a visually-based, lexical search interpretation will be examined.

<u>Two Lexical Searches as an Alternative to a Phonological Option</u>

Assuming that word recognition always proceeded by a purely visual strategy, <u>all</u> phonological specifications for words would be lexically mediat-

ed so that any effect of phonological bivalence should have been restricted to pseudowords. By a word-specific strategy where response latency for words depends on finding a visual match for a particular entry, the phonologically bivalent nature of a visual array of characters should have been irrelevant. Clearly, in the present word recognition studies, subjects never employ a pure (single lexicon) visual strategy, but rather they engage a strategy that is sensitive to phonological or at least alphabetic ambiguity. Experimental manipulations that affect words and pseudowords differentially are generally interpreted to indicate the involvement of a word-specific strategy. In the present experiment, the ambiguity by lexicality interaction indicated that the degree of detriment due to ambiguity for words and pseudowords differed. Nevertheless, the stronger effect on words introduced by phonologically ambiguous letter strings is consistent with the original bivalent experiment (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978), where the effect of bivalence was significant only for words. It is possible, therefore, that the effect of phonological bivalence originates with the problem of matching holistic letter string patterns with particular lexical entries and that word recognition in Cyrillic and Roman requires two distinct (visually-defined) lexicons.

The general pattern of results for lexical decision and naming were impressively similar and the correlation between tasks supported the claim of the participation of a common knowledge structure in both tasks. (Note that for a correlation, the naming lexicon and the lexical decision lexicon need not be identical, they only need to be organized in the same way.) One possibility is that this correlation reflects a lexical contribution and that the phonological effect occurs because the letter string matches with two entries in the lexicon. The standard interpretation of this correlation between tasks (Forster & Chambers, 1973; Frederiksen & Kroll, 1976) is that it reflects a visually-defined search on some non-segmented letter pattern that is specific to real words. In the present experiment, however, this systematicity extended to pseudowords. Because pseudowords do not have particular entries in a visually-defined lexicon, the effect could not be visual or holistic and specific to particular words (or morphemes), unless one supposed that pseudowords as well as words can be described by the Roman and Cyrillic lexicons. Allowing that a response could follow immediately when an entry was identified or that multiple decisions could arise, different degrees of impairment to performance for ambiguous words and pseudowords could be expected and these modifications will be considered.

1. Alphabet-governed Lexical Searches: Terminating

Respecting the assumption that phonologically bivalent strings are slower because they entail a parallel visual search of two lexicons (one for Roman forms and one for Cyrillic forms), there are two possible ways in which the different visual alphabets are searched: If they are searched in parallel such that operating in two files slows all latencies, then performance on words composed entirely of shared letters should be impaired, regardless of whether the shared letters correspond to the same phoneme in each alphabet (common letters), e.g., JAJE read as /jaje/, or correspond to different phonemes in Roman than in Cyrillic (ambiguous letters), e.g., KACA read as /kasa/ or as /katsa/. In fact, words containing only common letters are no slower than words that contain letters unique to one alphabet (Lukatela et al., 1980; Feldman et al., 1981).

Alternatively, the alphabet files may be searched in a successive fashion. Actually, Lukatela, Savić, Gligorijević, Ognjenović, and Turvey, (1978) have refuted an account of phonological bivalence based on two serial visual alphabet searches, because lexical decision to bivalent strings that were words by either alphabet reading, e.g., KACA (so that search would be successful in either alphabet file) was no faster than to strings that were words in Roman and pseudowords in Cyrillic, e.g., KOBAC. Likewise, pseudowords composed exclusively of common letters, e.g., TAKA (so that they have the same phonological reading in both Roman and Cyrillic) were no slower than pseudowords that contained letters unique to one alphabet (Lukatela et al., 1980). In sum, accounts of this detriment based on successive visual searches of two lexicons would predict that the presence of letters shared by the Roman and Cyrillic alphabets, regardless of their common or ambiguous phonemic value, should influence recognition, but this result was not observed.

Alternatively, perhaps only letter strings containing both ambiguous and common (and no unique) characters foster two alphabet searches. While the distinction between ambiguous and common letters shared by the two alphabets is phonological rather than visual, this option is worthy of consideration here because it encompasses both words and pseudowords and it treats bivalence as the result of complications in lexical search. If the probability of beginning search in either alphabet is equal, then for an ambiguous Cyrillic word (which is a Roman pseudoword), search will start in the correct alphabet file one half of the time. On the average, the subject need search one and one half files to recognize an ambiguous word. To reject an ambiguous pseudoword (which is a pseudoword in both Roman and Cyrillic), however, two full alphabet files need be examined on every trial. This terminating search-based account would predict that phonologically bivalent letter strings should be slower when they are pseudowords than when they are words: Both alphabets always must be considered before a "no" response is possible. For words, however, sometimes the search will begin with the appropriate alphabet and responding will not be delayed. Counter to this prediction, in the present experiments, latencies for lexical decision on ambiguous words and pseudowords did not differ (while analogous latencies for the unequivocal alphabet transcription of the same strings did differ). Moreover, a visual search might predict a trade-off between errors and reaction time (at least for real words), and higher variances among reaction times for individual ambiguous words--where lexical search can terminate--than for ambiguous pseudowords--where lexical search is necessarily exhaustive of two lexicons. In the present experiment, however, these measures were positively correlated and an analysis of variance on errors produced the same general results as on reaction time. Finally, in the present experiments, counter to the predictions of any visual search based account, the effect of phonological bivalence assessed within forms of the same letter string was <u>greater</u> for lexical decision on words than on pseudowords. (The type x group x lexicality interaction was not significant for naming due to the high variability in the pseudoword data. Therefore, no comparison of word and pseudoword latencies in naming is offered.) One other modification to the two-visually-defined lexical search model will be considered because it can account for the relative degree of bivalence among words and pseudowords.

2. Alphabet-governed Lexical Searches: Nonterminating

The larger effect of phonological bivalence for words than for pseudowords in lexical decision invites the notion of competing responses: For words, subjects must decide between the "yes" response engendered by the Cyrillic reading and the "no" response engendered by the Roman reading. For pseudowords, however, both readings would necessitate a "no" response. Until now, it was assumed that lexical search terminated immediately when a lexical entry was selected (or equivalently, as Coltheart has suggested, that visual and phonological strategies did not operate at the same rate). If search does not terminate immediately, then response competition becomes a viable description (and further, responding by a phonological strategy such as pseudowords traditionally require need not be slower than responding by a visual strategy). In this case, the phonologically derived pseudoword reading could influence the lexical reading of an ambiguous letter string so that both a positive and a negative decision are indicated. In fact, this account could work for naming as well as for lexical decision. Remember that in naming, the detrimental effect of bivalence appeared greater for pseudowords. There, there were two acceptable articulations while for words, only one reading produced a word. (Instructions specified to read the letter string as a word if it could be read as such.)

In general, attributing the detriment due to phonological bivalence to interfering responses complements the claim (Shulman, Hornak, & Sanders, 1978) that phonological effects in English may reflect a supplemental storage medium to improve visually-based performance rather than the descriptors by which a word was recognized. Since the memory-based account suggests a contribution by the lexicon to this phonological effect, it would not explain why the evidence of a phonological storage should be so much more pronounced in Serbo-Croatian than in English. More important, a nonterminating visual search of two alphabetically-defined lexicons and its consequence, a lexically-derived description of bivalence, cannot account for one crucial aspect of the present data.

In the present experiments, the detriment incurred by phonologically bivalent letter strings varied as a function of the number and distribution of ambiguous characters. Counter to any visually-defined search account of word recognition, these phonological results were exaggerated for words relative to pseudowords and were more stable in lexical decision, where there was no correlation between word frequency and degree of impairment, than in naming. In general, the degree of impairment increased with number of ambiguous characters, and two ambiguous characters within one syllable were more difficult than two ambiguous characters in different syllables. As an alternative to a visually defined search, if the nature of the Serbo-Croatian orthography and the general effect of phonological bivalence are reconsidered in terms of procedural knowledge or pattern analyzing operations (Kolers, 1975a), then perhaps this effect can be better captured in phonologically analytic rather than purely visual terms.

Recognition Strategies in Serbo-Croatian: A Phonological Priority

By tradition, visual strategies are presumed to be word-specific and are not appropriate for pseudowords. But the appropriateness of different stra-

tegies for word and pseudoword recognition must be the outcome of, not the starting point, for a description of lexical knowledge. Therefore, it is important to note that even if words and pseudowords were described by common lexical predicates, so that a visual, word-specific strategy is in principle possible, the effect of phonological bivalence cannot be rationalized by searches through a visual lexicon of even two lexicons. In the later experiment by Lukatela (Lukatela et al., 1980), as in the present experiment, an effect of bivalence was obtained for pseudowords. Because it slowed words and pseudowords in the same way and was independent of the number of lexical readings for each letter string, those investigators (Lukatela et al., 1980) proposed an account of the detriment due to phonolgical bivalence that was independent of word-specific knowledge and was based on the rate at which a description of the letter string that was appropriate for lexical search could be derived. This is reminiscent of a pattern analyzing procedure (Kolers, 1975a, 1975b, 1976) in that the systematic variability in word recognition is captured by the operations to apprehend visual patterns, rather than a search among substantive knowledge structures such as the lexicon model usually implies. Given the nature of the Serbo-Croatian language and the systematic relation between orthography and phonology, the present results suggest a pattern analysis for word recognition that proceeds in terms of the phonology, is independent of the lexicon, and is sensitive to component orthographic structure.

The pattern of results for lexical decision and naming was remarkably similar and the consistently high correlations between lexical decision and naming suggested that a common knowledge operation proceeded for all types of words and pseudowords in both tasks. Traditionally, this correlation has been interpreted as implicating the lexicon, a visually-defined word (or morpheme) specific knowledge structure, but in the present experiments, this correlation obtained for pseudowords as well as for words. In general, the major results demonstrated a very robust effect of phonological bivalence, and any account of this effect in terms of visual search of lexical structure, even one that allowed the inclusion of pseudowords, proved incomplete to encompass the significant effect of the distribution of ambiguous characters. In sum, there was no reason to conclude that the bases for lexical decision and for naming diverged: Both tasks entailed a phonological strategy even when it actually hindered performance.

GENERAL DISCUSSION

In the word recognition studies conducted in English, the phonological word-nonspecific strategy is often characterized as optional while the visual word-specific strategy is characterized as mandatory. The possibility of two strategies should actually diminish any phonological effect since, at least in English, the visual strategy is purported to operate faster than a phonological strategy (Coltheart et al., 1977). Nevertheless, the present experiment on Serbo-Croatian provided no evidence favoring this claim. On the same grounds, larger phonological effects (or weaker lexical effects) would be expected for the naming of words than for lexical decision to words, but this was not confirmed. In addition, the subjects in the present experiments all learned Cyrillic as their first alphabet and there is evidence that this early experience governs facility with the alphabets, even in mature readers

(Lukatela, Savić, Ognjenović, & Turvey, 1978). In the present experiments, all the ambiguous strings that were words, were words in their Cyrillic reading. If subjects had an option of employing a word-specific strategy exclusively, then in these experimental conditions it would have been optimal to reduce the availability of the (Roman) pseudoword reading and engage only the (Cyrillic) word reading. Nevertheless, these readers could not eliminate a phonological strategy in word recognition even when it was obviously detrimental to performance. In sum, the magnitude of the effect of phonological bivalence for words and pseudowords suggests that for skilled readers of Serbo-Croatian, the phonological strategy is neither slower nor optional.

Phonologically bivalent letter strings retarded performance relative to the unique alphabet transcription of the same form and this has been interpreted as evidence of a phonological strategy in word recognition. The question of a lexical contribution to the specification of phonology has not been resolved, however. Although there is evidence of a phonological strategy that is sensitive to sub-morphemic component structure, this does not eliminate the possibility of exploiting morpheme or word units, that is, a lexical specification of other aspects of phonology. Nevertheless, no currently available visually defined word-specific search model has proven adequate because that class of model proceeds holistically and is not phonographically analytic. In this discussion, no consideration of a lexical contribution that works concurrently with a lexically independent contribution has been delineated, and yet there is no reason why a lexicon-independent and a lexicon-derived phonological specification could not be implicated if, ultimately, the magnitude of the detriment due to phonological bivalence depends, among other factors, on the lexical status of the alternate reading.

In the word recognition literature, there has been a tendency to treat all aspects of knowledge about words in terms of substantive knowledge and to assume that the connection between newly presented words and previously acquired knowledge about words entails a search and match procedure in the internal lexicon. Issues in current theories of reading and word recognition focus on whether this match occurs in terms of predicates that reference visual aspects or predicates that reference phonological aspects of the written word. For alphabetic orthographies in general and for the shallow orthography of Serbo-Croatian in particular, these predicate types are not easily distinguished. Instead, in the present experiments, the distinction between strategies has been recast in terms of a contrast between holistic word-specific and phonologically analytic word-nonspecific strategies where the focus of a word-specific strategy is the word or morpheme and the focus of the word-nonspecific strategy is the phoneme. It was concluded that naming and lexical decision for both words and pseudowords are non-optionally phonologically analytic.

The dominant theories of reading and word recognition have been developed in English and have assimilated the idiosyncracies of this phonologically deep orthography into the theory. Comparisons with Serbo-Croatian, with its phonologically shallow orthography, invites the differentiation of the universal aspects of this particular theory of reading from the language-specific contribution. In the literature on word recognition based on English, it is often claimed that the acquisition of reading skill entails a shift away from a phonological recognition strategy (LaBerge & Samuels, 1974; Frederiksen,

1981) and, given that the English orthography references morphology as well as phonology, this may be true. By contrast, the written form of Serbo-Croatian has preserved a consistent reference to phonology and the character of this orthography is evident in the present studies of word recognition among skilled readers. Unlike reading in English that demonstrates a priority for a visual strategy, skilled reading in Serbo-Croatian retains a phonological priority.

REFERENCE NOTE

1. Feldman, L. B. The influence of lexical status on a phonological strategy for word recognition in Serbo-Croatian. Manuscript in preparation, 1981.

REFERENCES

Baron, J. Mechanisms for pronouncing printed words: Use and acquisition. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977, 175-216.

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 386-393.

Barron, R. W. Reading skill and phonological coding in lexical access. In M. M. Gruneberg, R. N. Sykes, & D. E. Morris (Eds.), Practical apects of memory. London: Academic Press, 1978.

Bauer, D. W., & Stanovich, K. E. Lexical access and the spelling-to-sound regularity effect. Memory & Cognition, 1980, 8(5), 424-432.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Coltheart, M. Lexical access in simple reading task. In G. Underwood (Ed.), Strategies of information processing. London: Academic Press, 1978.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. Phonological encoding in the lexical decision task. Quarterly Journal of Experimental Psychology, 1979, 31, 489-502.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI. London: Academic Press, 1977.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. Phonological coding and lexical access. Memory & Cognition, 1978, 6, 391-402.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, this volume.

Forster, K., & Chambers, S. M. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 627-635.

Frederiksen, J. R. Sources of process interactions in reading. In C. A. Perfetti & A. Lesgold (Eds.), Interactive processes in reading. Hillsdale, N.J.: Erlbaum, 1981.

Frederiksen, J. R., & Kroll, J. F. Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 361-379.

Gelb, I. *A study of writing*. Chicago: University of Chicago Press, 1952.

Gibson, E. J., & Levin, H. *The psychology of reading*. Cambridge, Mass.: M.I.T. Press, 1975.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, *75*, 554-570.

Gibson, E. J., Shurcliff, A., & Yonas, A. Utilization of spelling patterns by deaf and hearing subjects. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Gleitman, L., & Rozin, P. The structure and acquisition of reading 1: Relations between orthographies and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Towards a psychology of reading*. Hillsdale, N.J.: Erlbaum Associates, 1977.

Glushko, R. J. The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 674-691.

Glushko, R. J. Principles for pronouncing print: The psychology of phonology. In C. A. Perfetti & A. Lesgold (Eds.), *Interactive processes in reading*. Hillsdale, N.J.: Erlbaum, 1981.

Goodman, K. S. Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association, 1976.

Gough, P. One second of reading. In J. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationship between speech and reading*. Cambridge, Mass.: M.I.T. Press, 1972.

Gough, P. B., & Cotsky, M. J. One second of reading again. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1977.

Henderson, L. Word recognition. In N. S. Sutherland (Ed.), *Tutorial essays in psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Kalić, D. *Akusticka analiza vokala u ekavskom govoru Srpskohrvatskog jezika: I Tehnicke mogucnosti prepoznavanja pomocu filtara*. (Acoustic analysis of vowels in Ekavian dialect of Serbo-Croatian: And technical possibilities of recognition by filters.) Unpublished doctoral dissertation, Faculty of Electrical Engineering, University of Belgrade, 1964.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In C. A. Perfetti & A. Lesgold (Eds.), *Interactive processes in reading*. Hillside, N.J.: Erlbaum, 1981.

Kolers, P. A. Three stages of reading. In H. Levin & J. Williams (Eds.), *Basic studies in reading*. New York: Basic Books, 1970.

Kolers, P. A. Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *3*, 689-701. (a)

Kolers, P. A. Specificity of operations in sentence recognition. *Cognitive Psychology*, 1975, *7*, 289-306. (b)

Kolers, P. A. Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 554-565.

Kolers, P. A. Reading and knowing. *Canadian Journal of Psychology*, 1979, *33*, 106-117.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing. *Cognitive Psychology*, 1974, *6*, 293-323.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. Language and Speech, 1978, 21, 142-165.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. On the relation between processing the Roman and Cyrillic alphabets: A preliminary analysis with bi-alphabetic readers. Language and Speech, 1978, 21, 113-141.

Magner, T. F., & Matejka, L. Word accent in Serbo-Croatian. University Park, Pa.: Pennsylvania State University Press, 1971.

Mattingly, I. G. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationship between speech and reading. Cambridge, Mass.: M.I.T. Press, 1972.

McQuade, D. V. Variable reliance on phonological information in visual word recognition. Language & Speech, 1981.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Functions of graphemic and phonemic codes in visual word-recognition. Memory & Cognition, 1974, 2, 309-321.

Rubenstein, H., Garfield, L., & Millikan, J. Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic reading in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-654.

Rubin, G. S., Becker, C. A., & Freeman, R. H. Morphological structure and its effect on visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 757-767.

Shulman, H. G., Hornak, R., & Sanders, E. The effect of graphemic, phonetic and semantic relationships on access to lexical structures. Memory & Cognition, 1978, 6, 115-123.

Smith, F. Understanding reading. New York: Holt, Rirehart, & Winston, 1971.

Spoehr, K. T. & Smith, E. E. The role of syllables in perceptual processing. Cognitive Psychology, 1973, 5, 71-89.

Stanovich, K. E., & Bauer, D. W. Experiments on the spelling-to-sound regularity effect in word recognition. Memory & Cognition, 1978, 6, 410-415.

Taft, M. Recognition of affixed words and the word frequency effect. Memory & Cognition, 1979, 7, 263-272.

Taft, M., & Forster, K. I. Lexical storage and retrieval of prefixed words. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 638-647.

Venezky, R. L. Regularity in reading and spelling. In H. Levin & J. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.

Venezky, R. L. The structure of English orthography The Hague: Mouton, 1970.

FOOTNOTES

[1]There are exceptions to this characterization: For example the "d" in predsednik is generally interpreted as /t/. The number of violations is small, however.

[2]Two aspects of vowel accent (tone: rising/falling, length: long/short) are not captured by the written form. While vowel accent may differentiate between two semantic interpretations, this distinction is often ignored expecially in the dialects of the larger cities (Magner & Matejka, 1971). Moreover, vowel identity, at least as it is defined by formant structure in some restricted phonemic environments, is not distorted by variations in accent (Kalić, 1964).

WORD RECOGNITION WITH MIXED-ALPHABET FORMS

Laurie Beth Feldman and Aleksandar Kostić+

Abstract. In order to assess the influence of visually distorted print on word recognition, subjects named two styles of visually-distorted Serbo-Croatian words. Each word was repeated in several different distorted versions. For one group, the visual distortion entailed mixing characters from the Roman and Cyrillic alphabets (e.g., КИFЛА). While both of these alphabets are generally used to transcribe Serbo-Croatian, they are never mixed within a word. For the other group, the visual distortion entailed mixing case (e.g., KiFla). On the first trial, latencies to name words written in PURE Roman form (e.g., KIFLA) were no faster than latencies to name mixed alphabet forms. In addition, after training, mixed case forms were slower to name than mixed alphabet forms. It was concluded that for Serbo-Croatian, mixed alphabet visual distortions do not impair performance on a word naming task, but that mixed case distortions may not always function as mixed alphabet does. The assumption that word recognition is based on a familiar visual form was called into question.

While there is considerable debate about the role of phonology in the identification of real words in English, it is usually assumed that the familiar visual form of the word facilitates lexical access in studies of reading and word recognition. By definition, the graphemic characters of an alphabetic writing system correspond (approximately) to phonemes. Therefore, the distinction between a visual, word-specific strategy and a phonologically analytic strategy hinges on a (word-nonspecific) linguistic analysis of orthographic structure and on its consequence: an appreciation of the contribution of phonology to the formation of a visual pattern of alphabetic characters. Some researchers (Coltheart, Besner, Jonasson, & Davelaar, 1979) have claimed that the phonological strategy is optional and can be suppressed when it impedes performance, but that the visual strategy is always mandatory. In order to eliminate the value of a visual recognition strategy experimentally, visually distorted letter strings are presented for recognition in lexical decision or naming tasks. In the work with English, this distortion is commonly introduced by alternating upper- and lowercase letters within a word (e.g., Pollatsek, Well, & Schindler, 1975; Baron & Strawson, 1976; Mason,

---

+Also University of Connecticut and University of Belgrade.
Acknowledgment. Special thanks to Michael Turvey, Georgije Lukatela, and Predrag Ognjenović. We would also like to thank Ignatius Mattingly for insightful comments on an earlier version of this manuscript.
This research was supported in part by NICHD Grant HD 08495 to the University of Belgrade, in part by NICHD Grant HD 01994 to Haskins Laboratories, and in part by NSF dissertation fellowship BNS 7924409 to Laurie Feldman.

1978), and any disruption of linguistic information designated by case (e.g., proper noun, sentence initial word) has been ignored.

Although a phonological strategy depends on the analysis of the component orthographic structure in order to apprehend the phonology, it is generally assumed that the disruption incurred by alternating upper- and lowercase letters affects a visual word-specific strategy more than a phonological word-nonspecific strategy. The underlying assumption is that a visual word-specific strategy exploits overall visual shape of a word or transgraphemic features and avoids a phonological analysis. In contrast, a phonologically analytic strategy is insensitive to holistic (and letter features of) visual form and focuses on graphemic units. By this reasoning, effects of visual distortion on word recognition are traditionally interpreted as evidence of a visual, word-specific recognition strategy.

As described elsewhere (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978; Lukatela & Turvey, 1980), the relation between the written and spoken forms of Serbo-Croatian differs from the relation between the written and spoken forms of English in several respects. Essential to the present investigation, Serbo-Croatian is written in two different alphabets, Roman and Cyrillic. Although a small set of characters overlap, most of the characters are unique to one alphabet or the other (see Figure 1 and Table 1). (Also, see Feldman, 1981, this volume, for a more complete description). In addition, Serbo-Croatian has a shallow orthography with a relatively simple mapping between grapheme and phoneme so that it is not necessary to consider the morphological (or orthographic) context in which a particular grapheme occurs in order to assign it a phonemic value (see Feldman, 1981, this volume, for a more complete discussion). As a result, the orthographic conditions of Serbo-Croatian permit an unusual mixed-alphabet type of visual distortion of letter strings without interfering with the phonological interpretation of particular sequences of graphemes or introducing linguistically misleading case alternations.

It should be noted that although all readers in Yugoslavia learn to read both Roman and Cyrillic at an early age, the two alphabets seldom appear together in text. Most certainly, the characters of the two alphabets will never appear mixed within a word. By writing words in a combination of Roman and Cyrillic characters, an unprecedented visual word form can be generated.

In the present experiment, words printed in a mix of orthographically unique Cyrillic and Roman characters were presented in a naming task to one group of subjects. Another group of subjects named mixed case Roman forms of those same words. Onset to vocalization for the two styles of visual distortion was compared across trials. Even if subjects are initially no slower with bi-alphabetic patterns than with pure alphabet letter strings, repeated practice at analyzing distorted print may facilitate performance over trials. When new test words are presented on a subsequent trial, the visual pattern analyzing skill can then be extended to new and different letter strings so that latencies will not be prolonged. However, mixed alphabet forms may prove to be qualitatively different from mixed case forms. If the alternation of uppercase and lowercase characters poses a special problem, e.g., a linguistically anomolous situation, then mixed case and mixed alphabet forms may both be facilitated over trials, but these two distortions may

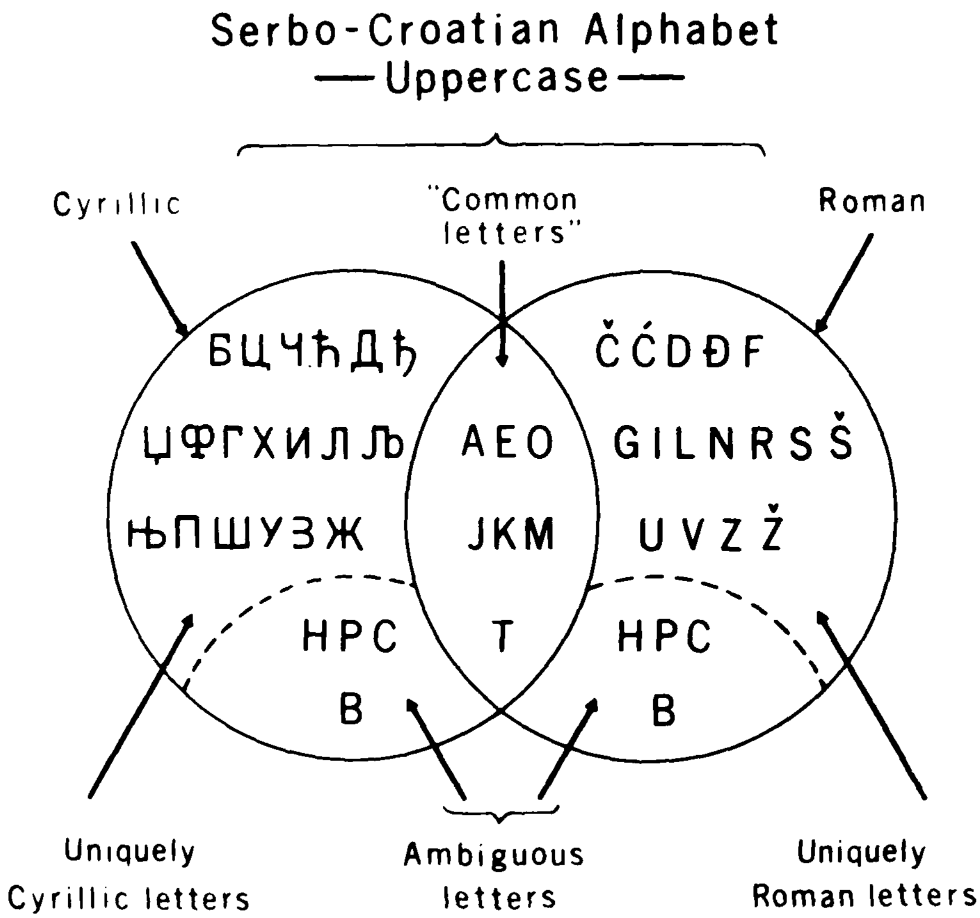


Figure 1. Letters of the Roman and Cyrillic alphabets.

TABLE 1

SERBO-CROATIAN				
ROMAN		CYRILLIC		
PRINTED		PRINTED		LETTER NAME IN I.P.A.
UPPER CASE	LOWER CASE	UPPER CASE	LOWER CASE	
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	nə
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

function differently when applied to new test words. Since there is substantial evidence that skilled reading of Serbo-Croatian as assessed by both the lexical decision and naming tasks is necessarily phonological (see Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela et al., 1978; Feldman, 1981), if totally unfamiliar mixed-alphabet visual distortions do not impair performance relative to pure alphabet forms, then perhaps it is the visual word-specific strategy that is optional in Serbo-Croatian.

METHOD

Subjects

Thirty-six first year students of psychology at the University of Belgrade participated in this study in partial fulfillment of course requirements. Thirty-one of those subjects had learned to read Cyrillic first, and the five subjects who had learned Roman first were approximately equally distributed between conditions.

Stimuli

All stimuli in the experiment were words containing between four and six letters. Words were selected in pairs so as to be phonologically similar in that each pair had at least three letters in common and each pair began with the same letter (e.g., ULAZ - UZDA, KIFLA - KUGLA). No words contained characters shared by both the Roman and Cyrillic alphabet (e.g., P, H, B, C) so that they commanded a different phonological interpretation, depending on alphabet.

In the CASE condition, words were presented in a mix of upper- and lowercase Roman letters. Each training word was presented five times in different configurations of alternating case (e.g., KiFla, kJfLa, KifLA). In the ALPHABET condition, the same words was presented in a mix of Roman and Cyrillic uppercase letters. As in the case condition, each training word was presented five times in different combinations of alternating alphabet (e.g., КИFЛA, КИФLA, KIФLA). Following training, a new set of ten test words was presented in the same style of distortion. The set of ten test words and the set of ten practice words were balanced for frequency and word length. Preceding the session for each of the two experimental groups, two practice items, one of which was in a PURE Roman (non-alternating form), and one of which was in the appropriate distorted form, were presented. In summary, each subject viewed 62 slides, which included two practice words, five repetitions of each of ten training words, and one presentation of each of ten test words.

Procedure

Subjects were required to name each word aloud as quickly as possible. The experimental session was divided into six (consecutive) trials, each consisting of one presentation of the same ten items (word order varied within each of the five training trials). For the ALPHABET condition, trials one through five contained randomized presentations of the same ten training words in different mixed alphabet configurations, while trial six consisted of a new

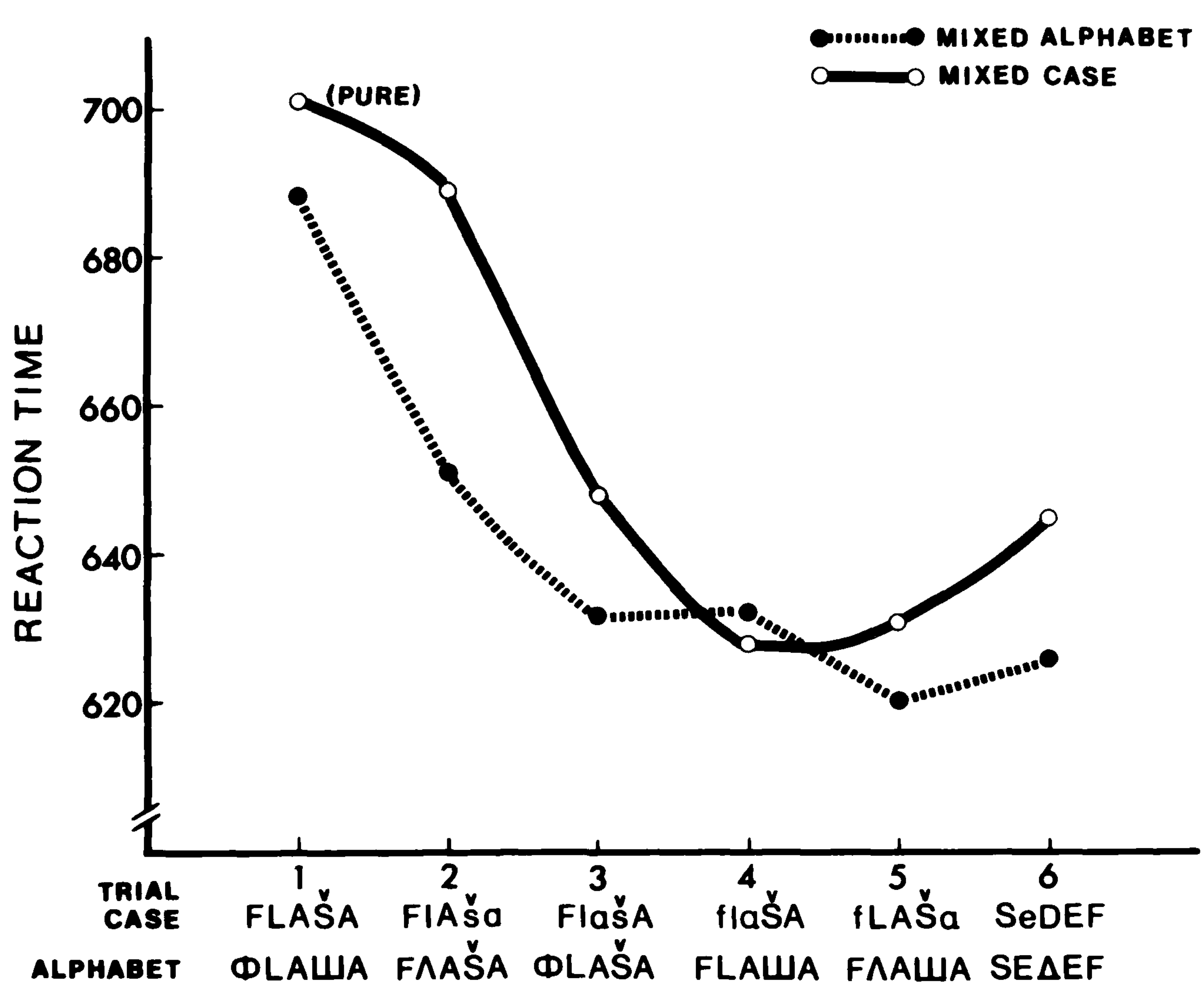


Figure 2. Mean reaction time to name training and test words for each trial.

set of ten test words, also in mixed-alphabet form. For the CASE condition, trial one consisted of the same ten training words in PURE Roman uppercase print. Trials two through five consisted of mixed case alternations of the same ten training words, while in trial six, a new set of the (same) test words was presented, again in a mixed case form. All stimuli were typed on Prima U Film (with the Cyrillic and Roman typeface closely matched for size and form) and were presented for 750 msec in a Scientific Prototype model GB tachistoscope. Reaction times were measured from the onset of the visual display to the onset of vocalization by a voice operated relay.

In summary, each of two groups of subjects saw five versions of each of ten training words (selected in pairs for phonological similarity) in either distorted CASE or distorted ALPHABET form. On the sixth trial, ten new test words were presented in either a mixed case or a mixed alphabet form, consistent with the previous trials. Subjects had to name each letter string as quickly and as clearly as possible. Errors and reaction times were recorded. Practice items were not included in the analysis. In all, there were 60 reaction time measurements for each subject.

RESULTS AND DISCUSSION

An analysis of variance on correct responses with minimum and maximum latencies set at 300 msec and 1000 msec was performed. (By setting the reaction time limits at these levels, a total of 17 and 21 responses were eliminated from the ALPHABET and CASE conditions, respectively. These responses clustered on the affricate č/tʃ/. Given the variability among initial phonemes in both the practice and test words, this restricted distribution of errors suggests that the voice key was not adequately sensitive to the onset of that particular acoustic pattern.) Total errors were extremely low—3 and 1, respectively, and these were incorrect articulations.

The analysis revealed a significant decrease in naming latency over Trials $F(5,170) = 28.50$, $MS_e = 910.1$, $p < .01$, and no significant difference between mixed CASE and mixed ALPHABET conditions, $F(1,34) = .61$. Although the Case by Trial interaction missed significance, $F(5,170) = 2.03$, $MS_e = 910.1$, $p < .10$, t-tests were performed. Examination of the first presentation of each training word (Trial One) revealed no difference between mixed alphabet and pure Roman forms, $t(17) = 1.35$. Indeed, if anything, the mixed alphabet forms tended to be faster (see Figure 2). Examination of the first presentation of each test word (Trial Six) revealed a significant difference between mixed alphabet and mixed case distortions, $t(17) = 1.84$, $p < .05$. Generally in the data, there is a suggestion that across trials, mixed alphabet forms appeared slightly faster than mixed case forms, but only in Trial Six is a comparison of mixed case and mixed alphabet unconfounded with number of previous repetitions of the same word possible. There, mixed alphabet forms are named faster than mixed case forms.

Unfortunately, due to the design of the experiment, no direct comparison of pure Roman and mixed case forms was possible. Therefore, discussion of the "locus" of the detriment due to case distortion in the course of lexical access and recognition as delineated for English is not relevant (e.g., Pollatsek et al., 1975; Bauer & Stanovich, 1980). Perhaps any differ-

ence between mixed case and mixed alphabet distortions is better interpreted as evidence that not all visual distortions are equivalent in terms of the pattern analysis required for recognition (Kolers, 1976, 1979). In particular, as mentioned above, case alternations may signal linguistic information in a way that alphabet alternations do not. In that event, mixed alphabet forms provide a purer style of visual distortion.

Although such letter strings never occur in conventional print, t-tests of the means of the two conditions in Trial One revealed that mixed alphabet forms were no slower to name than pure Roman versions of the same set of words. This result violates all theories of word recognition that grant priority to the familiarity of some holistic visual form of the word. Some theorists (Baron, 1977; Coltheart et al., 1979) have claimed that in English the knowledge base for a naming task need not be synonomous with the knowledge base for lexical decision, and when the relation between the written and spoken form is particularly reliable, such as in the phonologically shallow orthography of Serbo-Croatian, this criticism is perhaps more forceful. Nevertheless, there is evidence to the contrary (Feldman, 1981, this volume). In that experiment, both the analogous pattern of reaction times (and errors) for lexical decision and naming as well as the correlation between latencies for individual words in the two tasks showed that subjects employ the same strategies in both the lexical decision and naming tasks. Additionally, there is already some evidence from a lexical decision experiment by Katz and Feldman (1981) that mixed alphabet forms are not consistently slower than pure word controls. In contrast to that experiment, where half of the control items were in pure Roman print and half of the items were in pure Cyrillic print so that both alphabets need be available in both the control and mixed alphabet experimental conditions, in the present experiment all pure alphabet forms were written in Roman. In sum, attempts to model word recognition in Serbo-Croatian with visual descriptors sometimes succeed by positing two different alphabet spaces or lexicons for Roman and Cyrillic, but no visual model could accommodate the two alphabets into a visually defined lexicon such as the present data on mixed alphabet forms require.

It has been suggested elsewhere that the special properties of a writing system may influence word recognition in particular languages. As mentioned above, the English orthography is not fully consistent in its mapping between written form and surface phonetic form and, to the extent that the phonetic form entails the phonemic form, this may be offered as a justification for the claim that in English a visual strategy is mandatory while a phonological strategy is optional (Goodman, 1976). By contrast, Serbo-Croatian is very reliable in the relation between written and spoken form. It has been demonstrated previously that in Serbo-Croatian a phonological strategy is not optional (Lukatela et al., 1978; Lukatela et al., 1980; Feldman, 1981, this volume). The results of the present experiment complement that claim: Distortions to visual form do not generally impair performance in a word recognition task. If such distortions selectively impair one strategy, then it must be the visual word-specific strategy that is optional in Serbo-Croatian.

REFERENCES

Baron, J. Mechanisms for pronouncing printed words: Use and acquisition. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977, 175-216.

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 386-393.

Bauer, D. W., & Stanovich, K. E. Lexical access and the spelling-to-sound regularity effect. Memory & Cognition, 1980, 8, 424-432.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. Phonological encoding in the lexical decision task. Quarterly Journal of Experimental Psychology, 1979, 31, 489-502.

Feldman, L. B. Visual word recognition in Serbo-Croatian is necessarily phonological. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, this volume.

Goodman, K. S. Reading: A psycholinguistics guessing game. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Del.: International Reading Association, 1976.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In A. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading. Hillsdale, N.J.: Erlbaum, 1981.

Kolers, P. A. Reading a year later. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 554-565.

Kolers, P. A. Reading and knowing. Canadian Journal of Psychology, 1979, 33, 106-117.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. Language and Speech, 1978, 21, 142-165.

Lukatela, G., & Turvey, M. T. Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading and dyslexia. Baltimore, Md.: University Park Press, 1980.

Mason, M. From print to sound in mature readers as a function of reader ability and two forms of orthographic regularity. Memory & Cognition, 1978, 6, 568-581.

Pollatsek, A., Well, A. D., & Schindler, R. M. Familiarity affects visual processing of words. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 328-338.

INTRA- VERSUS INTER-LANGUAGE STROOP EFFECTS IN TWO TYPES OF WRITING SYSTEMS*

Sheng-Ping Fang,+ Ovid J. L. Tzeng,++ and Liz Alva+++

Abstract. The relation between word processing strategy and the orthographic structure of a written language was explored in the present study. Three experiments were conducted using Chinese-English, Spanish-English, and Japanese-English bilinguals, respectively. Each subject was asked to perform a modified Stroop color-naming task where the stimulus and the response language were either the same or different. The magnitude of Stroop effect was greater in the intra-language condition than in the inter-language condition. When the magnitude of reduction of Stroop interference from the intra- to the inter-language condition was compared across all bilingual groups, an inverse relationship was found between the magnitude of reduction and the degree of similarity between the orthographic structures of the two written languages. It is concluded that reading logographic and phonologic symbols entails different processing mechanisms and that controversial issues in bilingual processing cannot be resolved without taking into account the effect of orthographic variations on the information processing system.

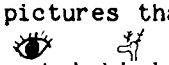
The invention of written symbols to represent spoken language is undoubtedly one of the most important achievements in the history of mankind. The written symbol has enabled us to overcome the limitations of space and time imposed by oral communication and has allowed us to extend our thoughts across centuries as well as continents.

There have been many different types of writing systems invented to represent various types of spoken languages. The designing principles for writing systems can be divided into two different categories. The first type of orthography evolved from the earlier semasiography, which expresses a

---

*This paper is to appear in Memory & Cognition, in press.
+University of California, Riverside.
++Also University of California, Riverside.
+++Abt Associates, Cambridge, Mass.
Acknowledgment. We thank Jessie S. H. Deou and Susan Stone for their great help in conducting the third experiment, and Katherine Malley for her helpful comments on earlier drafts. The final version of this paper has also benefited from comments made by Irvine Biederman and the reviewers. The present study was supported in part by an Inter-Campus Research Fund from the Academic Senate of the University of California, Riverside to the second author, and in part by a trainingship from the National Institute of Mental Health to the third author (supervised by Dr. Robert Singer).

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

general idea in picture drawings rather than a sequence of words in a sentence, to logographs with each symbol expressing a single particular morpheme. The concept underlying the development of this type of orthography is to map the written symbols directly onto words, from which meaning is generated. The second type of orthography evolved from the rebus (a representation of a word or phrase by pictures that suggest how a word is pronounced in the spoken language, e.g.,  for idea) to the syllabary and then to the alphabet. The concept behind it is sound writing. That is, the relation of sign to meaning is meant to be mediated through the sound system of the spoken language. This difference in how lexical units may be recovered from written symbols raises an important and interesting question: Do our visual information processing strategies differ when the information is presented in different formats? In recent years, this question has become of major concern among many cognitive psychologists (Biederman & Tsao, 1979; Gleitman & Rozin, 1977; Park & Arbuckle, 1977; Tzeng, Hung, & Garro, 1978).

That reading different writing systems may entail different information processing strategies is supported by some recent clinical and experimental observations. Sasanuma (1974) reported that the ability of Japanese aphasic patients to use logographic (kanji) and phonologic (kana) scripts can be selectively impaired. Parallel to this finding, in visual hemi-field experiments in which stimuli are presented to the right or left visual field briefly via a tachistoscope, a right visual field (i.e., left hemisphere) advantage is usually found for the recognition of phonologically based symbols such as English words or Japanese kana scripts, while a left visual field advantage is found for the recognition of single Chinese characters (Tzeng, Hung, Cotton, & Wang, 1979). Furthermore, in a cross-language study that investigated the effects of language (Chinese vs. English) and mode of stimulus presentation (visual vs. auditory), Turnage & McGinnies (1973) found that visual input facilitated the learning for Chinese subjects whereas auditory input produced superior recall performance for American subjects. All these results seem to point out that readers of different scripts may have developed different processing strategies in order to achieve efficient reading. It is of utmost importance for cognitive psychologists to find out at which level of information processing these differences due to orthographic variations occur.

A recent study of Biederman and Tsao (1979) shed light on the issue of the orthographic variations by using a Stroop (1935) interference paradigm. It is an established fact that in the Stroop color-word test, it requires more time to name a series of color patches when the patches are themselves incongruent color names (e.g., GREEN in red ink) than when the patches are simple colored rectangles. Biederman and Tsao (1979) found a greater interference effect for Chinese subjects in a Chinese version Stroop color-naming task than for American subjects in an English version. They attributed this difference to the possibility that there may be fundamental differences in the perceptual demands of reading Chinese and English. Since the perception of color and the direct accessing of meaning from a pattern's configuration are functions that have been assigned to the right hemisphere, it was suggested that during the Stroop test these two functions might be competing for the same perceptual capacity of the right hemisphere. This competition could have been avoided in the English Stroop test because reading English and naming color are executed by different hemispheric mechanisms. Biederman and Tsao further speculated that there may be some fundamental differences in the

obligatory processing of Chinese and English prints. They suggested that a reader of alphabetic writing cannot refrain from applying an abstract rule system to the word whereas a reader of Chinese may not be able to refrain from configurational processing of the logograph.

The conceptualization that reading different types of scripts automatically activates different types of perceptual constraints is an intriguing one. It leads to a unique prediction concerning the bilingual processing in a modified Stroop task. Suppose a Spanish-English bilingual subject is asked to name colors once in each of the two languages for color stimuli that are either Spanish color words, English color words, or control patches. Based on previous empirical findings (Dyer, 1971; Preston & Lambert, 1969), we can predict that color naming speed will be relatively slower when the naming language and the language of the color words are the same than when they are different. In other words, we can predict that the Stroop interference effect will be reduced in the inter-language condition as compared with that in the intra-language condition. But since both Spanish and English are alphabetic scripts that tend to activate similar obligatory processing strategies, the magnitude of reduction in the Stroop interference would not be much. Now suppose we ask a group of Chinese-English bilingual subjects to perform the inter- and intra-language Stroop tasks in which the interfering and the naming languages are either Chinese or English. It is again reasonable to predict that the inter-language condition will produce less Stroop interference than the intra-language condition. However, the most important question is whether the magnitude of reduction (from the intra- to the inter-language condition) will be greater, equivalent, or less for the Chinese-English bilinguals, as compared to that for the Spanish-English bilinguals. According to Biederman and Tsao's (1979) conjecture that reading alphabetic and logographic scripts entails different perceptual demands, one would predict that the magnitude of reduction (i.e., from the intra- to the inter-language condition) should be greater for the Chinese-English bilinguals than for the Spanish-English bilinguals. This expectation results from the assumption that while English and Spanish scripts activate similar obligatory processing strategies and thus are competing for the same perceptual demands, the Chinese and English scripts activate different obligatory processing strategies that do not interfere with each other. Experiments 1 and 2 were conducted to test this unique prediction generated from the considerations of orthographic variations and their relations to human information processing. Experiment 3 was conducted to further test this prediction while holding the phonological factor constant by using Japanese-English bilingual subjects.

METHOD

Experiment 1

Subjects. Thirty Chinese-English (C-E) bilinguals with normal color vision served as subjects. All were students at the University of California. Twenty of them were recruited from the Riverside campus and the remaining ten were from the Berkeley campus. All subjects had learned Chinese as their first language. All of them passed TOEFL (Test of English as a Foreign Language) before they were admitted into the University of California. Based upon their naming latencies of English and Chinese color terms (printed in black ink), all of them should be classified as Chinese dominant.

Materials. Three stimulus boards were prepared: one control board, one color-word board in English, and one color-word board in Chinese. Each board measured 40.6 x 50.8 cm^2.

The control board was constructed with six rows of ten 3 x 3 cm^2 patches, the colors of which were either red, blue, green, or brown. The patches were spaced 2 cm apart within each row and the rows were spaced 3 cm apart. Among the 60 patches, each of the four colors appeared 15 times in a random arrangement except that no color ever appeared twice in succession.

On the English board, the color arrangement was identical to that on the control board while each patch was replaced with an English word indicating an incongruent color name. Due to the physical nature of English words, each color word was 1.5 cm tall and up to 3 cm wide, centered in the place where the patch would have been. Words and colors used on this board were red, blue, green, and brown (Note: they are all monosyllabic words). Each word and color appeared 15 times randomly and no word or color appeared twice in succession.

The Chinese board resembled the English version in all aspects except that each English word was transformed into its corresponding Chinese character and measured 3 x 3 cm^2. The characters used on the Chinese board were 紅, 藍, 綠, and 褐, representing red, blue, green, and brown, respectively. The Chinese characters are monosyllabic in nature.

Design and Procedure. Each subject was given six tasks: (1) color naming of patches in English, (2) color naming of patches in Chinese, (3) color naming of English color-words in English, (4) color naming of English color-words in Chinese, (5) color naming of Chinese color-words in English, (6) color naming of Chinese color-words in Chinese. The order of administration was random.

Before the experiment started, the subject sat in front of a table while the stimulus board was placed on it, covered with a heavy blank paper sheet. The experimenter first explained the task and procedure to the subject. The subject was asked to perform each task as accurately and as quickly as possible, and to correct mistakes wherever possible. The subject was also asked not to point at the items while naming their colors. It was especially emphasized not to read the words but to name the colors of them instead. The subject was then asked to respond to two practice items, one Chinese character 黃 (representing yellow) in pink ink and another character 紫 (representing purple) in yellow ink. After proper responses were made, the experiment started. Each time a stimulus board was to be displayed, the subject was informed of the type of task to be performed. The stimulus board was covered again as soon as the task was completed. Color naming times for entire boards were recorded with a stopwatch to the nearest tenth of a second. Time between tasks was minimal, representing only the delay required to record data and obtain the new stimulus board.

Experiment 2

Subjects. Thirty Spanish-English (S-E) bilinguals with normal color vision served as subjects. All had learned Spanish as their first language with half of them Spanish dominant and the other half English dominant by their own estimates. However, based upon their naming latencies of English and Spanish color words (printed in black), all of them should be classified as Spanish dominant.

Materials. Three stimulus boards were used in Experiment 2, namely, one control board, one English color-word board, and one Spanish color-word board. Both the control board and the English board were identical to those used in Experiment 1. The Spanish board resembled its English counterpart in all aspects except that each English word was transformed into its Spanish equivalent. The Spanish equivalents were rojo, azul, verde, and cafe.

Design and Procedure. Each subject was given six tasks: (1) color naming of squares in English, (2) color naming of squares in Spanish, (3) color naming of English color-words in English, (4) color naming of English color-words in Spanish, (5) color naming of Spanish color-words in English, (6) color naming of Spanish color-words in Spanish. The order of administration was random. The instruction and procedure were the same as those in Experiment 1. Color naming times for entire boards were recorded with a stopwatch to the nearest tenth of a second.

RESULTS AND DISCUSSION

For each subject, the color naming time for the entire board was transformed into the naming time for a single item in milliseconds. This transformation procedure was applied to each of the six tasks and then the mean color-naming time for each of the six tasks was calculated based upon these transformed scores across the whole group. The data of the C-E bilinguals are presented in Table 1 (Experiment 1) and the data of the S-E bilinguals are presented in Table 2 (Experiment 2). Note that scores in parentheses represent the magnitude of the Stroop interference effect.

At first glance, the data presented in Table 1 seem to suggest that English color words produce greater Stroop interference (492 msec) than Chinese color characters (402 msec), a result at odds with that obtained by Biederman and Tsao (1979). However, careful reflection reveals that this comparison between our data and those of Biederman and Tsao may not be a valid one. In the present experiment, English is the second language for our subjects whereas in Biederman and Tsao's experiment, English is the native language for their American subjects. Thus, the data as shown in Table 1, should not be taken as an instance of failure to replicate Biederman and Tsao. In fact, our concern here is not to compare the degrees of interference between the Chinese Stroop task and the English Stroop task. Rather, the concern is with whether or not English and Spanish words (being both alphabetic scripts) would activate the same processing mechanism such that switching languages in a bilingual Stroop task should not reduce the amount of interference as much as in the case of switching between English and Chinese (a logographic script).

Table 1

Mean Color Naming Times (msec per item) for C-E Bilinguals on the Stroop Tasks (N = 30).

	English color-word	Chinese color-word	Control square	Mean
English Response	1431 (605)	1128 (302)	826	(454)
Chinese Response	1098 (378)	1221 (501)	728	(440)
Mean	(492)	(402)		

Note. Numbers in parentheses indicate the amount of interference (color-word minus control square).

Table 2

Mean Color Naming Times (msec per item) for S-E Bilinguals on the Stroop Tasks (N = 30).

	English color-word	Spanish color-word	Control square	Mean
English Response	1169 (495)	1017 (343)	674	(419)
Spanish Response	1166 (446)	1110 (398)	720	(418)
Mean	(470)	(366)		

Note. Numbers in parentheses indicate the amount of interference (color-word minus control square).

But before we examine the data pertinent to the above concern, let us clarify one particular point about the rationale behind the methodology. It can be argued that in no situation do subjects visually process words in the two languages simultaneously and that we may have a confusion between input (reading) and output (naming) mechanisms. Consequently, one may ask on what basis we can expect reading and naming to engage in one similar set of mechanisms. This question can be answered quite easily on empirical grounds. First, an automatic speech recoding of visually presented words is an established fact and it occurs in processing words written in alphabetic as well as non-alphabetic (such as Chinese, Japanese, etc.) scripts (Erickson, Mattingly, & Turvey, 1977; Tzeng, Hung, & Wang, 1977). Second, an automatic graphemic recoding of auditorily presented words has recently been established in a series of experiments by Seidenberg and Tanenhaus (1979) and by Nolan, Tanenhaus, and Seidenberg (1981). More importantly and interestingly, further studies on the graphemic recoding phenomenon by Tanenhaus, Flanigan, and Seidenberg (in press) demonstrated that such an automatic graphemic-recoding was responsible for slowing down color-naming responses in a Stroop-like task. Similar findings were also reported by Conrad (1978). Therefore, our assumption that the orthographic factor is involved in a color-naming task is completely justified.

Let us now examine the data presented in Tables 1 and 2 with respect to predictions made earlier in this paper. First of all, the Stroop interference effect was indeed reduced in the inter-language condition as compared with that in the intra-language condition. There was a 213 msec per item reduction for the C-E bilinguals and a 48 msec per item reduction for the S-E bilinguals. And indeed, the magnitude of reduction appeared greater for the former than for the latter.

A one-tailed planned comparison between inter- and intra-language Stroop effects was made for both bilingual groups. The magnitude of shift-language reduction was significant for the C-E subjects but not for the S-E subjects, $t(29) = 6.08$, $p < .0001$ and $t(29) = 1.48$, $p < .10$, respectively. Thus, the main prediction was confirmed. That is, the reduction scores of the two groups did differ significantly, and the magnitude of reduction was greater for the C-E bilinguals than for the S-E bilinguals.

For each bilingual group, a repeated-measures analysis of variance was also performed with the stimulus language as one factor and the response language as the second factor. For the C-E subjects, the main effect for the stimulus language was significant, $F(1,29) = 6.35$, $MSe = 38225$, $p < .05$, whereas the main effect for the response language was not, $F(1,29) < 1$. Also significant was the interaction between the two factors, $F(1,29) = 36.94$, $MSe = 36697$, $p < .001$. Further analysis of simple effects showed that there was significantly less interference whenever response and stimulus languages were different compared to the cases when they were the same. For the S-E subjects, the only significant effect found was the main effect of the stimulus language, $F(1,29) = 13.52$, $MSe = 24031$, $p < .001$, with English color-words resulting in greater interference than Spanish color-words in both response conditions.

For both S-E and C-E subjects, the stimulus language had much stronger control over the degree of interference effect as compared to the response

language. Both groups exhibited a significant main effect of the stimulus languages while, in both groups, response languages accounted for essentially zero percent of the total variance. These results suggest that the bilingual Stroop effect is more likely to be at the perceptual level than at the response level. The emphasis on the stimulus factor is in line with Biederman and Tsao's conjecture that the orthographic structure in the written language may play an important role in determining the magnitude of the Stroop effect. They also localize such an orthographic effect at the perceptual stage. They reason that different orthographic structures may impose different task demands such that different perceptual mechanisms are activated to meet these demands. This conceptualization also helps to explain the results of the two bilingual groups. Since both English and Spanish are alphabetic scripts, the perceptual mechanisms activated to process them are similar. Consequently, switching languages would not reduce the Stroop effect. On the other hand, Chinese logographs and English letters are two different scripts, and switching language means turning off one perceptual mechanism and turning on another one such that little interference would occur.

Based upon the above observations, we may induce a more generalized statement about the effect of the orthographic structure on the bilingual Stroop interference. That is, for any group of bilingual subjects, the magnitude of reduction from the intra- to the inter-language Stroop interference effect is a linearly decreasing function of the degree of similarity between the orthographic structures of the two languages. The validity of such an assertion can be tested by examining the patterns of the bilingual Stroop effects in the existing literature. To do this, we recalculated from the results of the present experiment and two other different bilingual experiments the magnitude of reduction of the Stroop interference from the intra- to the inter-language condition (Dyer, 1971, Experiment II, session 1; Preston & Lambert, 1969). All together, there were five types of bilingual subjects, namely, Chinese-English, Hungarian-English, Spanish-English, German-English, and French-English bilinguals. Wherever more than one experiment was run with respect to a certain type of bilingual, data were combined for that bilingual condition. We ranked these reduction scores according to their magnitude and obtained the following results (Table 3): Chinese-English bilinguals revealed a reduction of 213 msec; Hungarian-English, 112 msec; Spanish-English, 68 msec; German-English, 36 msec; and French-English, 33 msec per item. The ordering of the last three categories is particularly revealing. Why should switching between Spanish and English produce a greater reduction of interference than that between French and English or between German and English? It is certainly not intuitively obvious why Spanish and English are more orthographically dissimilar than French and English (or German and English). However, if we examine the spellings of color terms across these languages, then the deviation of Spanish becomes immediately clear. For example, red, blue, green, and brown are translated and spelled as rot, blau, grün, and braun in German; as rouge, bleu, vert, and brun in French; but as rojo, azul, verde, and cafe, respectively, in Spanish. Clearly, with respect to the color terms used in all these studies, Spanish color terms are orthographically more dissimilar to English color terms than both French and German. Correspondingly, we also observed a greater reduction of Stroop interference. This pattern confirms our expectation that the magnitude of reduction is a negative function of the degree of similarity between the orthographic structures of the two written languages. In other

words, the greater the orthographic similarity between the two languages, the stronger the competition for the same processing mechanisms and thus the smaller the reduction of Stroop interference from the intra- to the inter-language condition.

Table 3

Mean Reduction of Stroop Interference (msec per item) from the Intra- to the Inter-language Condition for Six Types of Bilingual Subjects from the Present Study and Experiments by Dyer (1971) and Preston and Lambert (1969)

Chinese-English	213
Kanji-English	121
Hungarian-English	112
Hirakana-English	108
Spanish-English	68
German-English	36
French-English	33

[a]Data from Experiment 3.

However, since orthographic similarity is highly correlated with phonological similarity, an alternative explanation is to attribute the effect of switching language to the phonological factor instead of the orthographic factor. Even though these two explanations are not necessarily mutually exclusive, it is important to determine which factor (orthographic or phonological) contributes more to the reduction of the Stroop interference. Experiment 3 was conducted to weigh the importance of the orthographic factor while holding the phonological factor constant.

EXPERIMENT 3

To answer the question whether the orthographic difference alone can account for the lexical processing and consequently the differential shift-language effects observed in the last two experiments, Japanese-English bilingual subjects were tested in Experiment 3.

Japanese is unique in the sense that three different types of scripts are concurrently used to represent the spoken language. Among the three types of scripts, Chinese logographs, referred to as kanji, are generally used to write the content words. The other two kinds of scripts, which are referred to as

hirakana and katakana and are syllabic in nature, are used for writing grammatical particles and foreign words, respectively. Though these three types of scripts differ in their writing styles, the words written with any one of the scripts are read in exactly the same pronunciation. This unique aspect of Japanese writing enables us to vary the orthographic structures while holding the phonological factor constant.

In this experiment, color-words were written in either kanji, hirakana, or English. With respect to the script/speech relationship embedded in the orthographic structure of the writing system, the hirakana script as a sound-writing system bears closer relation to the English script than the kanji logograph does. Following the arguments advanced by Biederman and Tsao (1979), it is reasonable to assume that the hirakana and English scripts are more likely to share a common processing mechanism than the kanji and English scripts. Accordingly, if the orthographic factor alone can effectively account for the differential reduction scores observed in Experiments 1 and 2, then the magnitude of reduction (from the intra- to the inter-language condition) should be significantly greater for the kanji-English condition than for the hirakana-English condition. On the other hand, if the phonological factor plays a more important role, then little difference in the magnitude of reduction should be observed between the kanji-English and the hirakana-English condition. Of course, there is always the possibility that both factors may play determinant roles in the bilingual Stroop effect.

What about the direct comparison between the pure cases (i.e., no language switching) of kanji and hirakana conditions? Biederman and Tsao (1979) demonstrated that more Stroop-type interference occurred in logographic than in alphabetic scripts. However, their demonstration has been criticized on the grounds of a possible confounding by two very different subject populations (Tzeng et al., 1978). In the present experiment, with kanji and hirakana scripts as the experimental materials, we were able to draw subjects from the same population and assign them randomly to two different conditions. Any demonstrated effect of orthography on the magnitude of the Stroop interference, therefore, should not be attributed to the subject factor.

Method

Subjects. Fifty Japanese-English bilingual students with normal color vision served as subjects. They were all natives of Japan and had at least six years of formal training in English as a second language. Most of them were enrolled in the ESL (English as a Second Language) Extension program and had been in the U.S. for less than one year. Thirty-eight subjects were tested at the University of California, Riverside campus and the remaining twelve were tested at the University of California, Berkeley campus. Subjects at both campuses were randomly divided into two groups. Group 1 was exposed to color-words in kanji and English while Group 2 was exposed to color-words in hirakana and English.

Materials. Four stimulus boards were prepared: one control board, one color-word board in English, one color-word board in hirakana, and one color-word board in kanji. For the consistency of grammatical form in Japanese, the four colors and color-names used in this experiment were red, blue, green, and purple. Both the control board and the English board resembled those used in

Experiments 1 and 2 except that the color and the word brown were replaced with purple in all cases. The hirakana board resembled the English version in all aspects except that each English word was transformed into hirakana. The hirakana equivalents were あか (AKA), あお (AUO), みどり (MIDORI), and むらさき (MURASAKI), representing red, blue, green, and purple. Their kanji counterparts were 3x3 cm^2 large and were the characters 赤 (red), 青 (blue), 緑 (green), and 紫 (purple). The control board, the English board, and the kanji version composed the stimuli for Group 1. The control board, the English board, and the kana version composed the stimuli for Group 2.

Design and Procedure. Subjects were randomly divided into two groups. All subjects were asked to perform the following four tasks: (1) color naming of squares in English, (2) color naming of squares in Japanese, (3) color naming of English color-words in English, and (4) color naming of English color-words in Japanese. Two additional tasks were assigned to Group 1 subjects: (5) color naming of kanji in English, and (6) color naming of kanji in Japanese. Similarly, subjects in Group 2 were asked to perform two additional tasks: (5) color naming of hirakana in English, and (6) color naming of hirakana in Japanese. The order of administration was random within each group and yoked between groups. The instruction and procedures were the same as those in Experiments 1 and 2. Color naming times for entire boards were recorded with a stopwatch to the nearest tenth of a second.

Results and Discussion

Color naming times for the entire card board were again transformed into reaction times of naming a single item in milliseconds. Table 4 shows the mean reaction times required for performing the six tasks. The scores of the Stroop effect shown in parentheses were analyzed separately for Group 1 and Group 2.

The scores of Stroop interference obtained from Group 1 were subjected to a repeated two-way ANOVA that examined the effect of the stimulus language and that of the response language. Statistical analysis revealed that the main effect of the stimulus language is significant, $F(1,24) = 8.11$, $MSe = 20083$, $p < .01$, whereas the main effect of the response language was not, $F(1,24) = 3.03$, $MSe = 32514$. There was also a significant interaction effect between the stimulus and response languages, $F(1,24) = 13.67$, $MSe = 27016$, $p < .005$. Further analysis suggested that the interaction resulted mainly from kanji scripts being exceptionally interfering when subjects are naming in Japanese.

A similar ANOVA was carried out on data of Group 2 subjects. The statistical analyses revealed neither an effect of the stimulus language nor an effect of the response language, $F(1,24) = 3.11$, $MSe = 16795$, and $F(1,24) = 2.00$, $MSe = 44964$, respectively. However, there was a significant interaction between these two factors, $F(1,24) = 9.50$, $MSe = 30645$, $p < .01$. Post-hoc analysis of simple effects showed that when subjects were naming in English, English scripts interfered more than hirakana, $F(1,48) = 4.98$, $MSe = 9930$, $p < .05$, and when subjects were naming in Japanese, hirakana interfered more than English, $F(1,48) = 30.04$, $MSe = 9930$, $p < .005$. In the presence of hirakana, naming colors in Japanese was more difficult than in English, $F(1,48) = 9.49$, $MSe = 37804$, $p < .005$, while naming colors in one language was not more difficult than in the other when English words were presented, $F(1,48) < 1$.

Table 4

Mean Color Naming Times (msec per item) for Japanese-English Bilinguals on the Stroop Tasks

	Group 1 (N = 25)				Group 2 (N = 25)			
	Eng. color-word	Kanji color-word	Control square	Mean	Eng. color-word	Kana color-word	Control square	Mean
English	994	954	704		990	928	721	
Response	(290)	(250)		(270)	(269)	(207)		(238)
Japanese	913	1115	681		910	1064	689	
Response	(232)	(434)		(333)	(221)	(375)		(298)
Mean	(261)	(342)			(245)	(291)		

Note. Numbers in parenthese indicate the amount of interference (color-words minus control square).

Of particular concern is whether differences in the orthographic structure play a decisive role in the magnitude of Stroop interference in a mixed-language condition. A one-tailed planned comparison between the intra- and the inter-language condition was made for each of these two groups. The magnitude of shift-language reduction was highly significant for both groups. There was a 121 msec per item reduction for Group 1 (kanji), $\underline{t}$ (24) = 3.68, $\underline{p}$ < .005, and a 108 msec per item reduction for Group 2 (hirakana), $\underline{t}$ (24) = 3.08, $\underline{p}$ < .005. However, the reduction scores of the two groups did not differ significantly, even though the direction of the difference was consistent with our expectation, $\underline{t}$ (48) = .28, ns. Apparently, the phonological factors contribute more to the reduction of Stroop interference in the mixed-language condition than the orthographic factor does.

Another comparison was made between the two conditions where both stimulus and naming languages were Japanese. Shimamura and Hunt (Note 1) conducted a Stroop experiment with color-words written either in kana or in kanji (a within-subject factor). They found a higher Stroop effect for kanji than for kana script with Japanese subjects. In the present experiment, color naming in Japanese did appear more difficult for the kanji version than for the kana version (434 vs. 375). Again, the difference is in the right direction. However, the difference was not statistically significant, $\underline{t}$ (48) = .23, ns.

According to the above results, it does not seem that a strong explanation based upon variations in orthography has gained support in Experiment 3. Yet, the orthographic factor cannot be totally dismissed without some cautious comments. In all comparisons made between kanji and hirakana processing, the direction of differences exhibited an expected pattern but the differences failed to reach a statistically significant level. However, we have noted that similar studies carried out in other laboratories (Shimamura & Hunt, Note 1; Biederman, personal communication) with a more powerful design (within-subject instead of between-subject) and with other dependent measures (e.g., error rates)[1] did report significant differences. Therefore, we think the orthographic factor does play a role, but may not be as important as the phonological factor, in the bilingual Stroop experiment.

A criticism has always been raised against the comparison of kanji and kana symbols in the color naming task. For fluent readers of Japanese, the color terms they read in everyday life are usually expressed in kanji script and rarely in kana. Hence, the greater interference observed for the kanji script may be attributable to this familiarity factor. To counter such an argument, Shimamura and Hunt (Note 1) and Biederman (personal communication) presented further evidence showing that in a simple word naming experiment (naming words printed in black), color terms written in kana were actually named much faster than color terms written in kanji. Similar findings were reported by Feldman and Turvey (1980). So, although color terms are more frequently written in the kanji form and although kanji are more compact graphic representations of words in general, naming time was consistently less for the kana. Thus, familiarity seems not to be a major factor in this case.

GENERAL DISCUSSION

In recent years, reading research has become a significant interdisciplinary endeavor with contributions from such diverse fields as anthropology, artificial intelligence, cognitive psychology, educational psychology, linguistics, and neuropsychology. The present study tackles the issue of word processing from a cross-language perspective. Since the way a spoken language is represented graphemically varies from language to language, it is essential to find out whether such orthographic variations impose different processing requirements on readers of different written scripts. Two questions are of particular concern in the present study. First, would different processing mechanisms be activated in reading the logographic and the alphabetic scripts? Second, does the particular pair of languages that a bilingual individual knows have a specific effect on the degree of language overlap? For instance, should Chinese-English bilinguals be considered as qualitatively different from Spanish-English bilinguals with respect to their lexical representations?

The first question can be answered more or less in an affirmative manner. Indeed, the idea that reading logographic and phonologic symbols entails different cognitive strategies and processing mechanisms has been supported by studies concerning aphasia (Sasanuma, 1974), visual lateralization effects (Tzeng et al., 1979), quantity-comparison tasks (Besner & Coltheart, 1979), and serial recall (Turnage & McGinnies, 1973). Biederman and Tsao have suggested that there may be fundamental differences in the obligatory processing of alphabetic and logographic print. A reader of alphabet writing cannot refrain from applying an abstract rule system to the word, whereas a reader of Chinese cannot refrain from configurational processing of the logograph.

Answers to the second question are less unequivocal. On the one hand, we see that a rough estimate of the magnitude of reduction in the Stroop effect in mixed-language conditions (as compared to pure-language conditions) from among seven different types of bilingual subjects exhibits an orderly relationship between the orthographic structure and the amount of reduction. On the other hand, experiments with the two types of Japanese scripts only provide minimal support for the predictions generated from the consideration of orthography. Nevertheless, we also noted that data from other similar studies did provide much stronger support. Thus, we may conclude that the orthographic structure does play an important role, independent of the phonological factor, in the lexical formation of a bilingual subject.

The implication of such orthographic and phonological effects for research in bilingual processing is clear. We simply cannot, or should not, lump data of different types of bilingual subjects together and attempt to come up with a general statement about the processing mechanism. It has been the common practice of investigators of bilingualism to talk about L1 (first language) and L2 (second language) without paying much attention to the degree of orthographic and phonological similarities between the two languages. No wonder there is so much inconsistency from one bilingual study to another. For example, there is currently a controversy as to the pattern of the hemispheric dominance in L1 and L2 of a bilingual subject. It is conceivable that a Spanish-English bilingual should show a very different cerebral lateralization pattern from that of a Chinese-English bilingual (Tzeng et al., 1979). Thus, without taking into account the influence of the orthographic

structure, many controversial issues in bilingual processing are difficult to resolve.

The relation between language and thought has been a topic of intensive investigation for hundreds of years. Delineation of script/speech relationships and discovery of how the orthographic variations affect our information processing system will no doubt open up a new possibility for specifying the nature of symbol/thought interactions.

REFERENCE NOTE

1. Shimamura, A., & Hunt, E. Stroop interference tests with kanji and kana scripts. Unpublished paper, University of Washington, Seattle, 1978.

REFERENCES

Besner, D., & Coltheart, M. Ideographic and alphabetic processing in skilled reading of English. Neuropsychologia, 1979, 17, 467-472.

Biederman, I., & Tsao, Y. C. On processing Chinese ideographs and English words: Some implications from Stroop-test results. Cognitive Psychology, 1979, 11, 125-132.

Conrad, C. Some factors involved in the recognition of words. In J. W. Cotton & R. L. Klatzky (Eds.), Semantic factors in cognition. Hillsdale, N.J.: Erlbaum, 1978.

Dyer, F. N. Color-naming interferences in monolinguals and bilinguals. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 297-302.

Erickson, D., Mattingly, I., & Turvey, M. T. Phonetic activity and reading: An experiment with kanji. Language and Speech, 1977, 20, 384-403.

Feldman, L. B., & Turvey, M. T. Words written in Kana are named faster than the same words written in Kanji. Language and Speech, 1980, 23, 141-147.

Gleitman, L. R., & Rozin, P. The structure and acquisition of reading I: Relations between orthographies and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conferences. Hillsdale, N.J.: Erlbaum, 1977.

Nolan, S. D., Tanenhaus, M. K., & Seidenberg, M. S. Multiple code activation in word recognition: Evidence from rhyme monitoring. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 170-180.

Park, S., & Arbuckle, T. Y. Ideograms versus alphabets: Effects of script on memory in "Biscriptual" Korean subjects. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 631-642.

Preston, M. S., & Lambert, W. E. Interlingual interference in a bilingual version of the Stroop color-word task. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 295-301.

Sasanuma, S. Impairment of written language in Japanese aphasics kana versus kanji processing. Journal of Chinese Linguistics, 1974, 2, 141-157.

Seidenberg, M. S., & Tanenhaus, M. K. Orthographic effects on rhyme monitoring. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 546-554.

Stroop, J. R. Studies of interference in serial verbal reaction. Journal of Experimental Psychology, 1935, 18, 643-662.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. Orthographic and phonological activation in auditory and visual word recognition. Memory & Cognition, in press.

Turnage, T. W., & McGinnies, E. A cross-cultural comparison of the effects of presentation mode and meaningfulness in short-term recall. American Journal of Psychology, 1973, 86, 369-381.

Tzeng, O. J. L., Hung, D. L., Cotton, B., & Wang, W. S-Y. Visual lateralization effect in reading Chinese characters. Nature, 1979, 282, 499-501.

Tzeng, O. J. L., Hung, D. L., & Garro, L. Reading the Chinese characters: An information processing view. Journal of Chinese Linguistics, 1978, 6, 287-305.

Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.

FOOTNOTE

[1]Biederman also suggested that we examine the error rates across kanji and kana conditions. We did keep the records of errors in each condition. Because of the tremendous amount of individual differences and the uncertainty of the nature of these errors, we did not analyze them systematically. However, the overall pattern is consistent with the argument that the kanji Stroop task is much more difficult than the kana Stroop task. The mean errors committed in the kanji and kana conditions are 5.42 and 2.75, respectively.

CATEGORICAL PERCEPTION OF ENGLISH /r/ AND /l/ BY JAPANESE BILINGUALS

Kristine S. MacKain,+ Catherine T. Best,++ and Winifred Strange+++

Abstract. Categorical perception of a synthetic /r/-/l/ continuum was investigated with Japanese bilinguals at two levels of English language experience. The Inexperienced Japanese group, referred to as NOT Experienced, had had little or no previous training in English conversation. The Experienced Japanese had had intensive training in English conversation by native American-English speakers. The tasks used were absolute identification, AXB discrimination, and oddity discrimination. Results showed classic categorical perception by an American-English control group. The NOT Experienced Japanese showed near-chance performance on all tasks, with performance no better for stimuli that straddled the /r/-/l/ boundary than for stimuli that fell in either category. The Experienced Japanese group, however, perceived /r/ and /l/ categorically. Their identification performance did not differ from the American-English controls, but their overall performance levels on the discrimination tests were somewhat lower than for the Americans. We conclude that native Japanese adults learning English as a second language are capable of categorical perception of /r/ and /l/. Implications for perceptual training of phonemic contrasts are discussed.

Languages differ in their phonological and phonetic inventories. For example, in a particular language (L1), two phones may occur, while in another language (L2), the phones may not appear at all. Or, L1 and L2 may share two phones, but in L1 the phones may be phonologically contrastive, while in L2, they may occur in contextual or free variation rather than being used to distinguish meaning. Because of this variation across languages, several questions have been asked about the potential role of linguistic experience in the perception of phonological categories. Are speakers universally sensitive to the parameters that distinguish phonological contrasts in all languages, or does experience with the phonological categories of one's native language affect the perception of those contrasts? For native speakers of languages that do not make use of particular speech sounds in a phonological contrast,

+Cornell University Medical College.
++Also Teachers College, Columbia University.
+++University of Minnesota.

Acknowledgment. Supported by NICHD (NIH) postdoctoral fellowship grant #HD054707 to the first author, NINCDS (NIH) postdoctoral fellowship grant #NS05877 to the second author, NIH grant #HD01994 to Haskins Laboratories, and NIMH grant #MH21153 to James J. Jenkins, which supported the third author while she was a visiting scholar at Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

is the perception of those sounds affected? If so, can perception of a phonetic contrast be modified in adulthood through learning a language that does employ the contrast as a phonological opposition?

The first two questions have been answered to some extent by cross-language investigations of vowel and consonant perception. It has been found that linguistic experience with phonological contrasts does affect perception of them, at least for some vowels and consonants. For vowels, experience influences perceptual discrimination judgments made along an interval scale, but does not produce differential nominal judgments. Using nominal (same/different) judgments, Stevens, Liberman, Studdert-Kennedy, and Öhman (1969) found no difference in the ability of native Swedish and American-English speakers to detect differences in vowel contrasts that were phonologically distinct in Swedish but not in English. However, by employing a more sensitive interval scale discrimination measure, Terbeek (1977) found that language experience in monolinguals of five different languages does affect vowel perception. The perceptual distance between two vowels was judged to be much greater if the pair contrasted phonologically in the subjects' native language than if the pair was not a native contrast.

Linguistic experience also affects the location of phonetic perceptual boundaries between stop consonant contrasts. For instance, Voice Onset Time (VOT)--the time between release of articulatory closure and onset of phonation--is a sufficient cue for phonological categorization of stop consonants in perception (Lisker & Abramson, 1970) and production (Lisker & Abramson, 1967). These investigators found cross-language differences in the location of the perceptual boundary between "voiced" and "voiceless" phonetic categories along a synthetic stimulus continuum underlying VOT. For each language group, identification (Lisker & Abramson, 1970) and discrimination (Abramson & Lisker, 1970) responses were generally in close correspondence. Moreover, identification and discrimination responses for Thai and American-English speakers were different and generally corresponded to their respective stop voicing production distributions, reported in an earlier study (Lisker & Abramson, 1964). Similar effects of experience have been found with native Spanish speakers (Abramson & Lisker, 1973; Williams, 1977) whose VOT production distributions differ from both Thai and English. It appears, then, that experience with specific voicing contrasts among stop consonants determines the location of perceptual boundaries separating those phonological categories along the acoustic continuum.

The effects of linguistic experience just summarized suggest, in addition, the converse situation--that lack of experience with a given phonological contrast should result in a poorly-defined perceptual boundary separating the two members of that contrast. Cross-language studies on categorical perception of non-native phonetic contrasts have addressed this issue. Categorical perception is said to occur if the subject cannot discriminate speech sounds any better than she/he can identify them within different phonological categories. Under these conditions, equal increments along a phonetically relevant acoustic continuum are not discriminated unless the increment crosses the boundary between phonetic categories (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

In this vein, recent studies (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975; Mochizuki, Note 1) have assessed the perception of synthetic /r/-/l/ continua by native Japanese and native American-English speakers. Native Japanese speakers who have learned English as a second language in adulthood are notorious for having difficulty in discriminating /r/ from /l/. In spoken Japanese, the liquid /l/ does not occur. Although a form of /r/ ("rhotic") is said to occur phonemically, it fits the criteria of a flap [ɾ], and is more similar acoustically and articulatorily to the American-English voiced dental-alveolar flap [ɾ] than to the approximant [ɹ] in American English (Miyawaki, Note 2; Price, Note 3). The 21 Japanese subjects in the Miyawaki et al. (1975) experiment had all studied English for at least 10 years; however, their instruction did not stress conversational English and only two subjects had resided in an English-speaking country. The Japanese subjects completed an oddity discrimination task on a synthetic /r/-/l/ continuum that varied only the spectral configuration of the third oral formant (F3), considered to be the primary cue for the contrast in English. Presumably, neither endpoint corresponded to the spectral configuration of the Japanese /r/ category. American-English subjects completed both oddity discrimination and identification tasks. The latter showed typical categorical perception results; they divided the continuum consistently into two phonetic categories in the identification task, and discriminated between-category comparison pairs well but within-category comparison pairs poorly. In contrast, the Japanese did not discriminate the series categorically; discrimination was nearly random and was no better for comparisons that crossed the phonetic boundary than for those lying within either the /r/ or the /l/ category.

Whereas the Miyawaki et al. (1975) study included a test of /r/-/l/ discrimination by Japanese, Mochizuki (Note 1) tested nine Japanese speakers only on an identification test, which used a synthetic /r/-/l/ series (again, only the F3 spectral configuration was varied). Her Japanese subjects divided the continuum into two distinct phonetic categories with a perceptual boundary that closely corresponded to that of an American-English control group; this would seem at odds with the Miyawaki et al. report. However, it may be important that the English language experience of the Japanese in the two experiments differed somewhat. Although both sets of subjects had had similar levels of formal training with English, Mochizuki's subjects had all lived in an English-speaking country, and were still residing there at the time of testing (range = 6 months-4 years in U.S.).

The present investigation examined categorical perception of /r/ and /l/ by native Japanese speakers at two levels of English language experience, and compared their performance to that of native American-English speakers. The design was a replication and extension of Miyawaki et al. (1975) with the following changes: (1) The synthetic stimulus series included variation in both spectral and temporal acoustic dimensions that differentiate natural American-English /r/ and /l/ (Dalston, 1975). These redundantly cued stimuli were used in order to optimize the Japanese subjects' opportunity to show perceptual differentiation of the /r/-/l/ contrast. (2) In addition to the oddity discrimination task used in previous studies, an AXB discrimination task was included. This task has lower memory demands and is thought to provide a better opportunity for detecting auditory differences. (3) An absolute identification task was included for computing predicted discrimina-

tion performance by the American and the two Japanese groups. These three tasks provide an extensive perceptual profile for the phonological contrast with stimuli that closely resemble natural speech exemplars of the phonological categories in American English. The primary question of interest was whether Japanese-English bilinguals with relatively intensive experience conversing with native English speakers would identify and discriminate /r/ and /l/ according to American-English categories, while Japanese with less English conversation experience would show less categorical /r/-/l/ perception.

METHOD

Subjects

The American group was comprised of ten undergraduates (5 males, 5 females) recruited through notices posted on campus bulletin boards at Yale University. We recruited Japanese adults from the Yale community by telephone; 12 agreed to participate (7 males and 5 females). All were Japanese natives who had moved to the U.S. as adults, except for one young woman who had moved at 15 years. They filled out a language-experience questionnaire prior to the experiment, and two subgroups were chosen on the basis of the amount and quality of their English conversation experience (see Table 1). The Experienced group contained five subjects (2 males, 3 females) who had had intensive English conversation training by native American-English speakers. The other seven (5 males, 2 females) were designated NOT Experienced, by contrast, because they had had little or no native English conversation training. All subjects reported normal hearing in both ears. Pay for participation was $3.25/hr.

Stimuli

A /rɑk/ -/lɑk/ ("rock" -"lock") series was generated on the OVE-IIIc synthesizer at Haskins Laboratories. The endpoint stimuli were traced from spectrograms of /rɑk/ and /lɑk/ utterances by an American male. Although the F3 initial steady-state and transition direction is a sufficient minimal cue for the perception of the initial /r/-/l/ contrast by Americans (O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957; Miyawaki et al., 1975), the stimulus series used here included variations not only in spectral characteristics of F3, but also in spectral characteristics of F2 and in temporal characteristics of F1. Figure 1 provides a schematic spectrographic representation of the stimuli. The series contained ten nearly-equal steps[1] of concurrent change for the F3 onset frequency (between 1477 Hz and 2594 Hz for /r/ and /l/, respectively), and for F3 frequency at the point of inflection (between 1067 Hz and 1207 Hz). There were five equal steps of F1 transition abruptness (between 21 ms and 49 ms) so that each F1 configuration occurred in two of the stimuli in the series. (See Appendix A for a detailed specification of stimulus parameters.)

Procedure

All subjects took part in three tests during a single session: 1) forced-choice identification, 2) AXB discrimination, and 3) oddity discrimina-

Table 1

American English conversation experience of the Experienced and the NOT Experienced Japanese subjects.

Factors:	A % day conversing in English since coming to U.S.(25, 50,75, or 100%)	B # hr/wk in instruction on English conversation by native speakers	C # mo. experience in English conversation with native speakers
Experienced Japanese:			
S 7 (♂)a	75%	8	48
S 8 (♀)b	75%	10	48
S 9 (♂)a	75%	8	18
S10 (♀)b	25%	10	18
S11 (♀)c	25%	4	6
	$\underline{X} = \overline{55\%}$	$\underline{X} = \overline{8}$	$\underline{X} = \overline{27.6}$
NOT Experienced Japanese:			
S 1 (♀)c	25%	3	5
S 2 (♂)d	25%	0	5
S 3 (♂)d	25%	3	2
S 4 (♂)d	25%	0	6
S 5 (♀)c	25%	0	18
S 6 (♂)d	25%	0	18
S12 (♂)d	50%	0	6
	$\underline{X} = \overline{28.6\%}$	$\underline{X} = \overline{.86}$	$\underline{X} = \overline{8.7}$

[a] graduate student
[b] undergraduate student
[c] homemaker
[d] postdoctoral associate

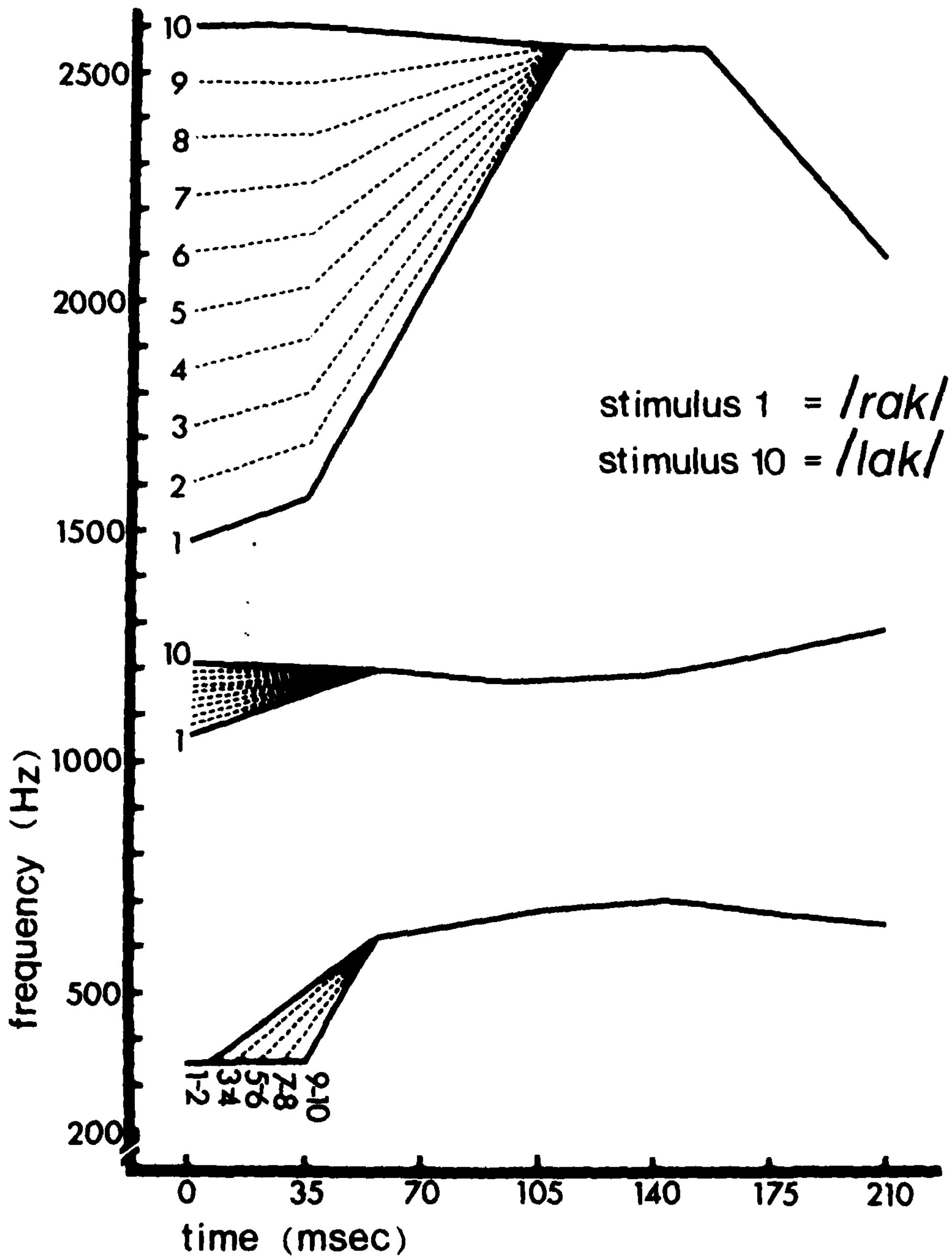


Figure 1. Schematic spectrogram representations of the ten stimuli in the synthetic /rɑk/-/lɑk/ series.

tion. Testing was conducted in a sound-attenuated chamber, with stimuli presented at a comfortable listening level (approximately 75 dB SPL) over TDH-39 headsets to groups of two to four subjects. The identification test consisted of 20 repetitions of the ten stimuli, randomized within each block of ten trials. Intertrial intervals (ITIs) were 2.5 seconds, and interblock intervals (IBIs) were 4 sec. Subjects wrote "R" or "L" for "rock" or "lock" on each trial, and chose the closer word for any ambiguous-sounding stimulus. These and subsequent instructions were typed in English for the Japanese subjects to read.

Subjects then completed an AXB discrimination test that contained ten repetitions of each of the two AXB orders for the seven possible 3-step stimulus pairings (1-4, 2-5, 3-6, 4-7, 5-8, 6-9, 7-10). Trials were blocked by 14 (2 orders x 7 AXB pairings), and were randomized within blocks. Within-trial interstimulus intervals (ISIs) were 1 sec, ITIs were 3 sec, and IBIs 6 sec. Subjects indicated for each trial whether the second item (X) matched the first (A) or third item (B).

Next, the subjects completed the oddity discrimination test, which contained eight blocks of 21 trials randomized across blocks of two. Each set of two blocks contained one each of the six oddity orders for the seven possible 3-step pairings. There were thus 24 trials for each of the comparison pairs. The subjects indicated whether the odd stimulus on each trial was first, second, or third.

RESULTS

Americans

The Americans showed classic categorical perception of /r/ and /l/ (Figure 2). Their identification responses (left-hand panel) showed a sharp category boundary near stimulus 5, and the endpoint stimuli (1 and 10) were identified with perfect consistency as /rɑk/ and /lɑk/, respectively.

Predicted discrimination functions were computed from the identification data, for both the AXB and oddity tests. For each discrimination test, distinct peaks in performance were obtained near the /r/-/l/ category boundary (center and right-hand panels, Figure 2). The data were analyzed by a two-way Stimulus Pairs (7 levels) X Functions (2 levels: obtained and predicted) analysis of variance (ANOVA). The Stimulus Pairs effect indicated that between-category performance peaks were higher than within-category performance on both the AXB test, $\underline{F}(6,54) = 11.45$, $p < .001$, and the oddity test, $\underline{F}(6,54) = 9.50$, $p < .001$. Obtained performance was somewhat better than predicted (solid vs. dotted lines, Figure 2), according to the Functions effect for both the AXB test, $\underline{F}(1,9) = 8.44$, $p < .025$, and the oddity test, $\underline{F}(1,9) = 5.25$, $p < .05$. However, post-hoc Tukey tests of pairwise comparisons (Glass & Stanley, 1970) revealed significant differences only for comparisons of clear-case stimuli against ambiguously-identified boundary stimuli (i.e., AXB pairs 2-5, 3-6, and 6-9; oddity pairs 3-6, 5-8, and 6-9). Obtained performance was no better than predicted for between-category comparisons (pair 4-7) and for clear within-category comparisons (1-4, 7-10). That is, obtained discrimination performance exceeded category-based predictions only

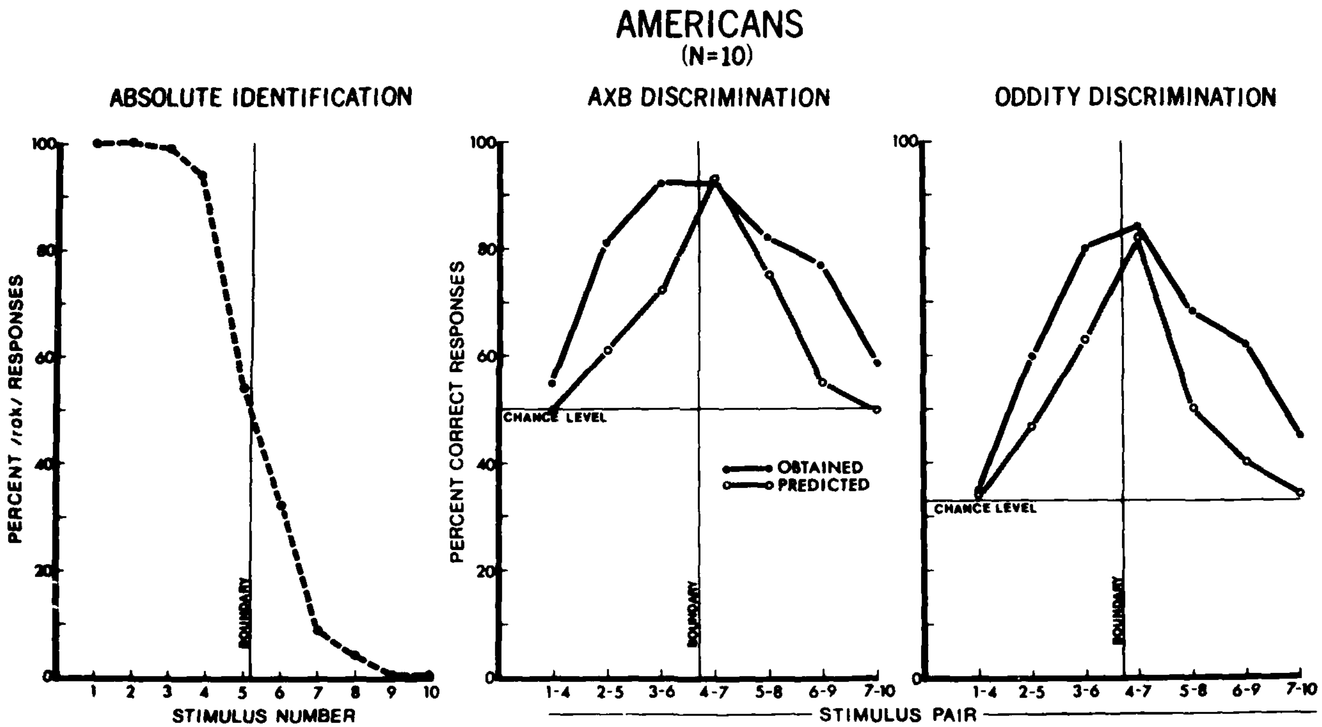


Figure 2. Response functions for the American group on the identification, AXB, and oddity tests.

when there were "category goodness" differences between stimuli within a phonetic category.

NOT Experienced Japanese

In striking contrast to the Americans, the identification data indicate poor /r/-/l/ classification by the Japanese with little English conversational experience (left-hand panel, Figure 3). Category judgments hovered near chance (50%) throughout the stimulus series, and even the endpoint stimuli were, on the average, only slightly differentiated perceptually (60% vs. 40% /rak/ responses).

As predicted by their identification results, the NOT Experienced Japanese performed little better than chance on the two discrimination tests. Although obtained performance on both tests (center and right-hand panels, Figure 3) appears to be slightly better than predicted, the Stimulus Pairs x Functions ANOVA on these data failed to show any significant differences. Thus, the data from this group replicate and extend the Miyawaki et al. (1975) results.

Experienced Japanese

While the results for the NOT Experienced group support the Miyawaki et al. (1975) suggestion that lack of experience with /r/-/l/ as a native phonological contrast limits the perception of that contrast, the data nonetheless pose some questions: Can the limitation in /r/-/l/ perception be overcome by adults, and if so, to what extent, and through what possible types of experience? The data for the Experienced Japanese (Figure 4) address these questions. All of these subjects had had intensive English conversation training with native American-English speakers and spent a larger percentage of their average day conversing in English than did the NOT Experienced Japanese (see Table 1). As shown in Figure 4, their identification data (left-hand panel, Figure 4) are quite similar to the American results, and contrast with those for the NOT Experienced Japanese.

In addition, the discrimination functions for these subjects, on both the AXB and oddity tests, were more similar to those of the Americans than those of the NOT Experienced Japanese (see group comparisons, Figure 5). Although their discrimination performance was not as high as that of the Americans, both discrimination tests revealed an increase in correct performance near the /r/-/l/ category boundary. The Stimulus Pair effect for the ANOVA on this group's discrimination data confirmed the significance of this discrimination peak on both the AXB test, $\underline{F}(6,24) = 3.981$, $p < .01$, and the oddity test, $\underline{F}(6,24) = 6.919$, $p < .001$. Unlike the Americans, however, obtained discrimination was not significantly better than predicted. The Stimulus Pairs X Function interaction for their AXB data, $\underline{F}(6,24) = 2.703$, $p < .05$, indicated that the between-category obtained function was significantly flatter than predicted (i.e., less distinct peak).

The contrasts and similarities in the identification functions for the three groups (shown in Figure 5) suggest that the occurrence and abruptness of an /r/-/l/ category boundary for the Experienced Japanese might be related to

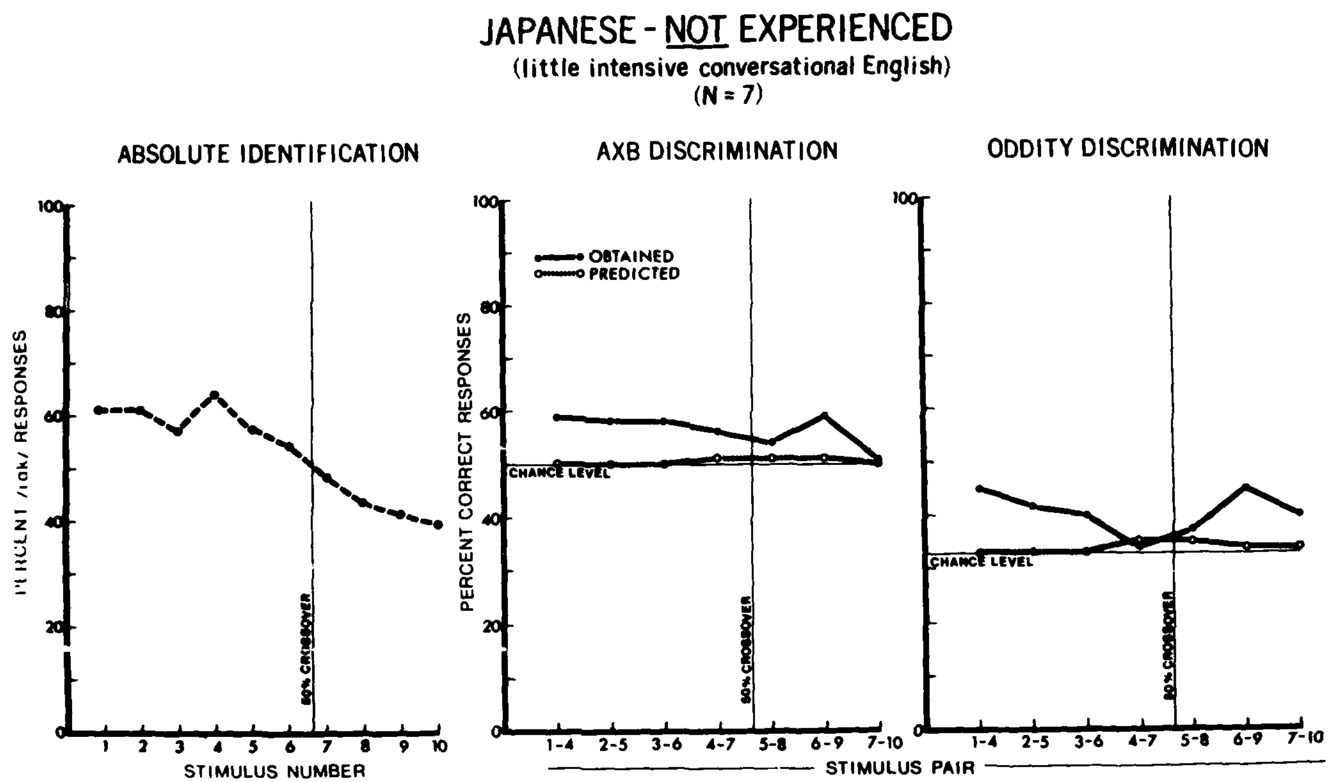


Figure 3. Response functions for the NOT Experienced Japanese group on the identification, AXB, and oddity tests.

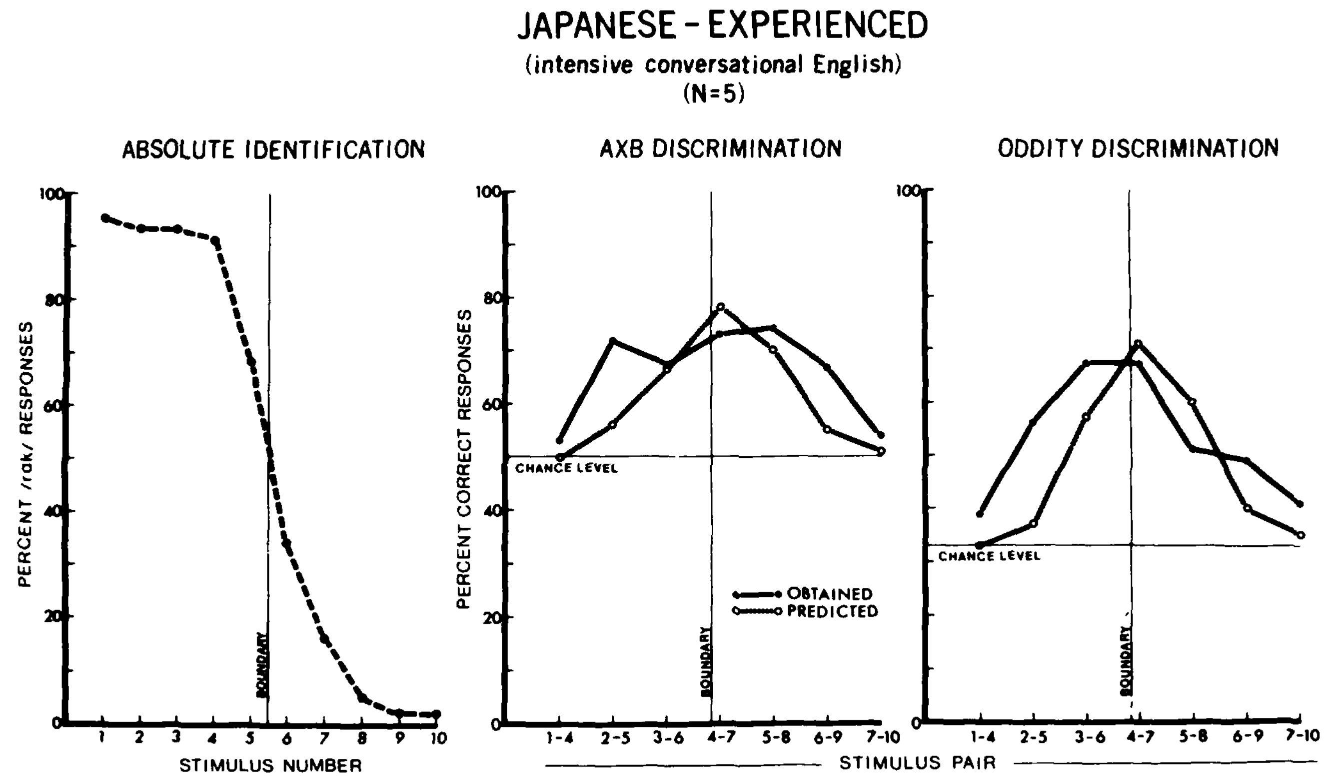


Figure 4. Response functions for the Experienced Japanese group on the identification, AXB, and oddity tests.

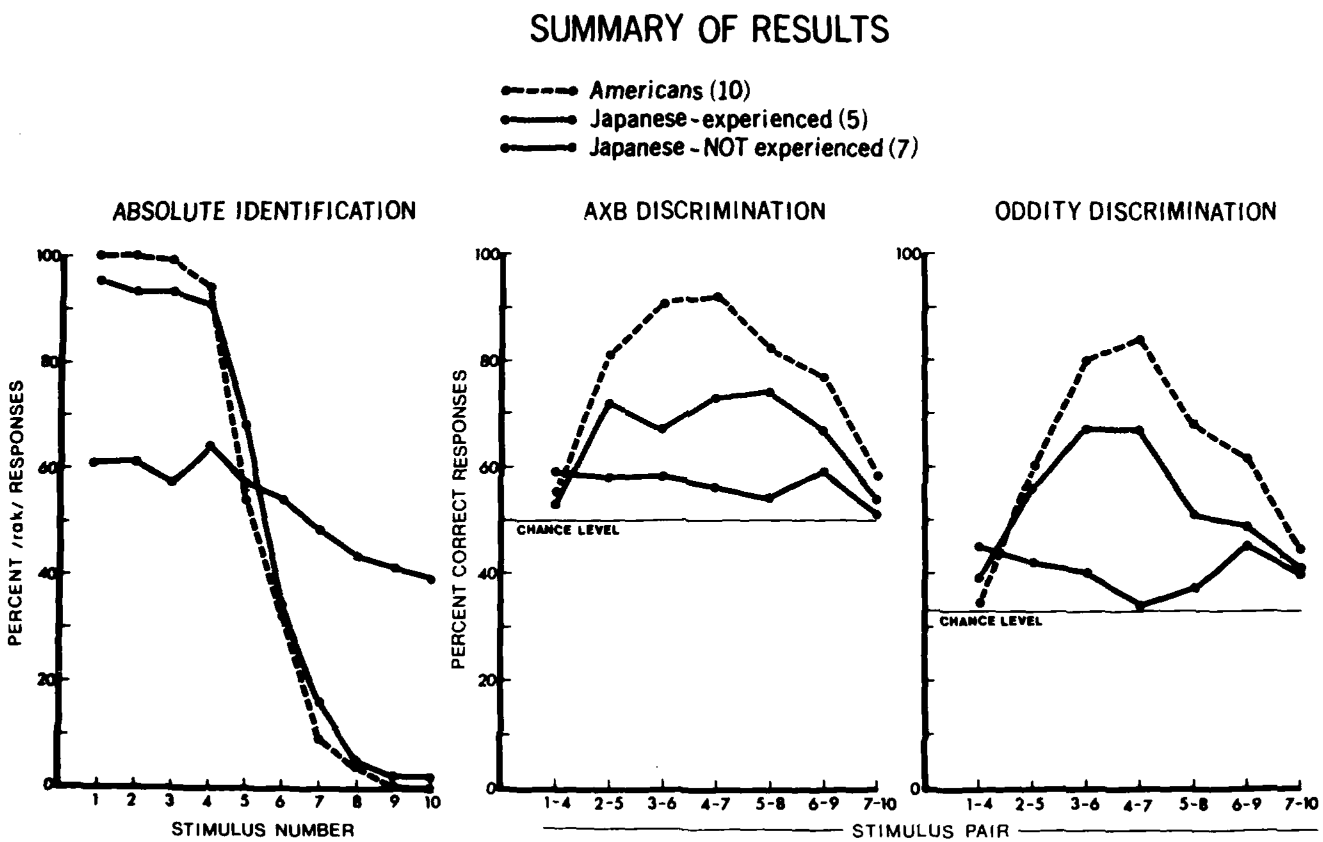


Figure 5. Comparison of Americans, Experienced Japanese, and NOT Experienced Japanese results on the three tests.

their greater conversational English experience, relative to the other Japanese group. To assess this possibility, a measure of the steepness of the category boundary was devised, for correlation with the English language experience factors listed in Table 1. For all individuals in each group, a narrow-range PROBIT analysis of the identification data was used to fit the best ogive to the 50% crossover region of the /rak/-/lak/ categorization function (see Figure 6). These analyses included the stimulus number closest to the individual's crossover, plus the adjacent higher- and lower-numbered stimuli. The ogives fit the data well--the χ^2 values failed to approach the 5.0 value (χ^2 range = 0.0-3.8) that would denote significant deviation between obtained data and fitted ogive at the .05 alpha level, with one minor exception for the least experienced subject in the Experienced group (S11: χ^2 = 5.21).

The slopes of these ogives were determined as a reflection of the abruptness and direction of the perceptual category change. Slope values range from a theoretical minimum of 0.0, a perfectly vertical shift from 100% to 0% /rak/ identifications, to a maximum of 1.0, an equally abrupt but phonetically inappropriate shift from 0% to 100% /rak/ responses. Very small slope values thus reflect a sharp and phonetically appropriate category boundary, whereas values at .5 represent a flat slope (no true boundary), and values greater than .5 would represent a phonetically incorrect category shift.

The boundary slopes for the Experienced Japanese were nearly as small (*M* = 0.038; range = 0.016 to 0.082) as for the Americans (*M* = 0.016; range = 0.01-0.025), while those for the NOT Experienced Japanese were noticeably larger (*M* = 0.346; range = 0.098 to 0.708). If there were a significant positive effect of English conversation experience upon the development of clear /r/-/l/ phonetic categories by the Japanese subjects, a strong negative correlation should be found between the boundary slope and the amount of experience. All three English experience factors listed in Table 1 showed a moderate-to-substantial negative correlation with boundary slopes, but Factor B (# hr/wk English conversation instruction by native speaker) showed the strongest negative correlation (*r* = -.67). Factor A (% day speaking English in U.S.) showed the smallest correlation (*r* = -.33), and the correlation for Factor C (# mo. experience speaking English with Americans) was -.41. Factor B, which is an indicant of the <u>intensity</u> of conversation instruction over an indeterminate period, was more strongly negatively correlated with boundary slopes than was even the total number of hours spent in English conversation instruction (# hr/wk X # wks instructed).

<u>An Anomaly: Subject M.K.</u>

After completion of the above data analysis, we had the opportunity to test an additional Japanese subject, whose English experience placed him in the NOT Experienced group. He was a newly-arrived postdoctoral associate at Yale, and had only been in the U.S. for two weeks at the time of testing. He spoke English less than 25% of the day, and had had no English conversation training by a native speaker, nor was he conversant in any other language besides Japanese. His performance on the three tests, surprisingly, was more similar in many respects to the Experienced Japanese than it was to the NOT Experienced group (see Figure 7). His identification function showed a sharp

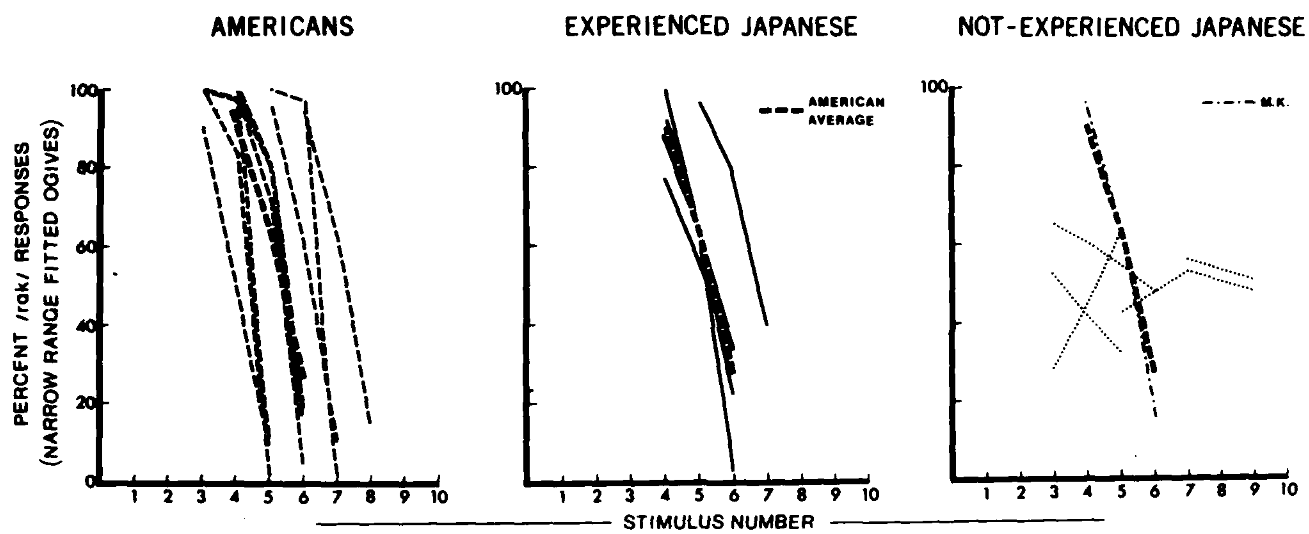


Figure 6. Narrow-range fitted ogives for the category boundaries given by the individual subjects in the three groups.

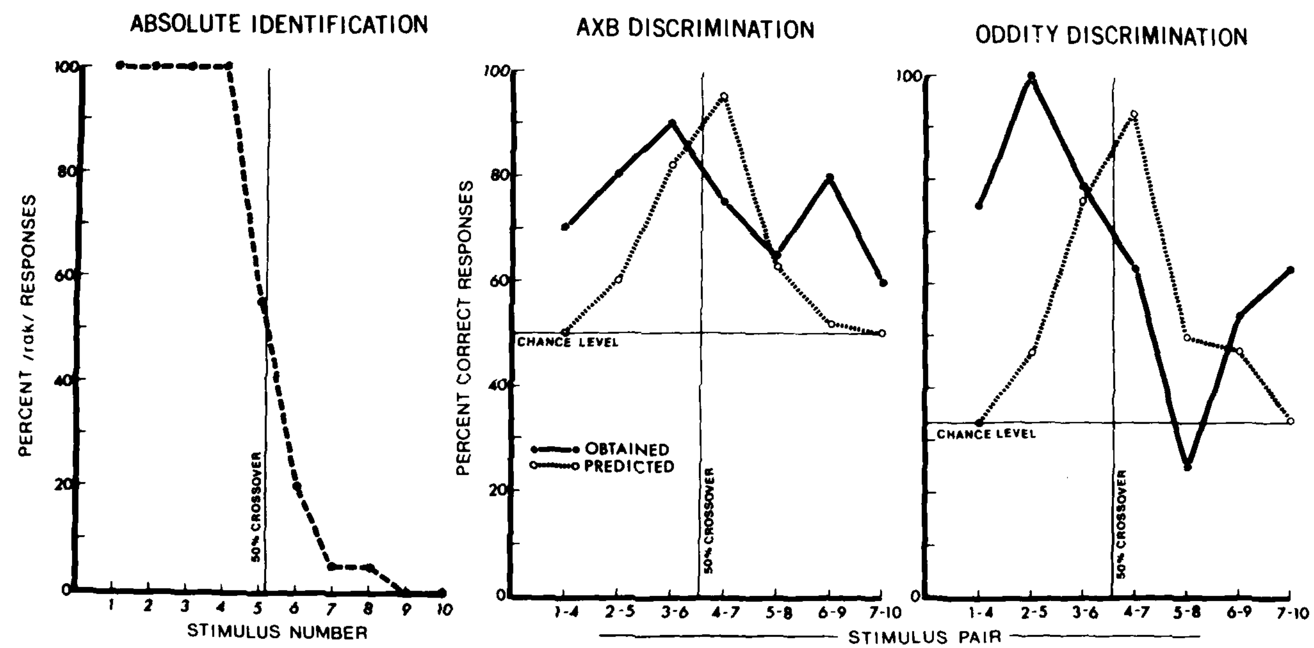


Figure 7. Response functions for subject M. K. on the identification, AXB, and oddity tests.

shift near the American boundary (intercept = 5.28), and his discrimination performance was higher even than several of the Experienced Japanese. The major distinction between his data and those of the Experienced Japanese was that both his discrimination functions were bimodal; neither of the peaks fell at his /r/-/l/ boundary, as would have been expected had his ability to discriminate the stimuli been limited in a direct way by his phonetic classifications of them, and as it was for the Americans and the Experienced Japanese.

Relative Performances on AXB vs. Oddity Tests

Both discrimination tests were included in our study because of claims that AXB comparisons are less demanding on memory than are oddity comparisons, and provide less of a bias toward phonetic categorization. It has been argued that these circumstances allow subjects to have better access to nonphonetic, pre-categorical stimulus information under AXB conditions than under oddity conditions (for fuller discussion of this, see Best, Morrongiello, & Robson, 1981 -- Experiment 2). These claims lead to the prediction that AXB performance will be better than oddity performance, especially for the NOT Experienced Japanese, since they could use nonphonetic auditory memory to aid performance on the AXB test more than on the oddity test. In addition, the oddity boundary-related peak should be sharper than the AXB boundary peak, especially for the Americans and probably for the Experienced Japanese. That is, an auditory memory-related improvement in AXB over oddity performance would affect the within-category judgments more than the between-category judgments.

In order to make a direct AXB-oddity performance comparison, it was necessary to adjust for the difference in chance level performance on the two tests (50% for AXB and 33.3% for oddity). Therefore, performance on the two discrimination tests was re-calculated as percentage of above-chance performance. These above-chance performance data were analyzed separately for each group in two-way Test (AXB vs. oddity) x Stimulus Pairs (1-4 through 7-10) ANOVAs.

As can be seen in Figure 8, AXB performance was better than oddity performance for the Americans, according to their significant Test effect, $F(1,4)=14.90$, $p < .05$. However, the Test x Stimulus Pairs effect for this group did not reach significance, suggesting that, contrary to the auditory memory/phonetic bias predictions, their oddity peak was not consistently sharper than their AXB peak. Their between-category performance was no less affected by the test format than was their within-category performance. Moreover, again in contradiction to the auditory memory/phonetic bias predictions, the Test effect and the Test x Stimulus Pairs interactions failed to reach significance either for the NOT Experienced Japanese, or for the Experienced Japanese. It is especially surprising that the oddity discrimination performance of this latter group is closer in form to the ideal picture of categorical discrimination than is their AXB function, in light of suggestions that the oddity paradigm biases subjects toward phonetic categorization rather than discrimination of auditory properties. That is, for these adults, who are learning a non-native contrast, a bias toward phonetic categorization (oddity test) leads to better discrimination of between-category comparisons than does a task with a presumably reduced bias toward phonetic categorization and a lower memory demand (AXB).

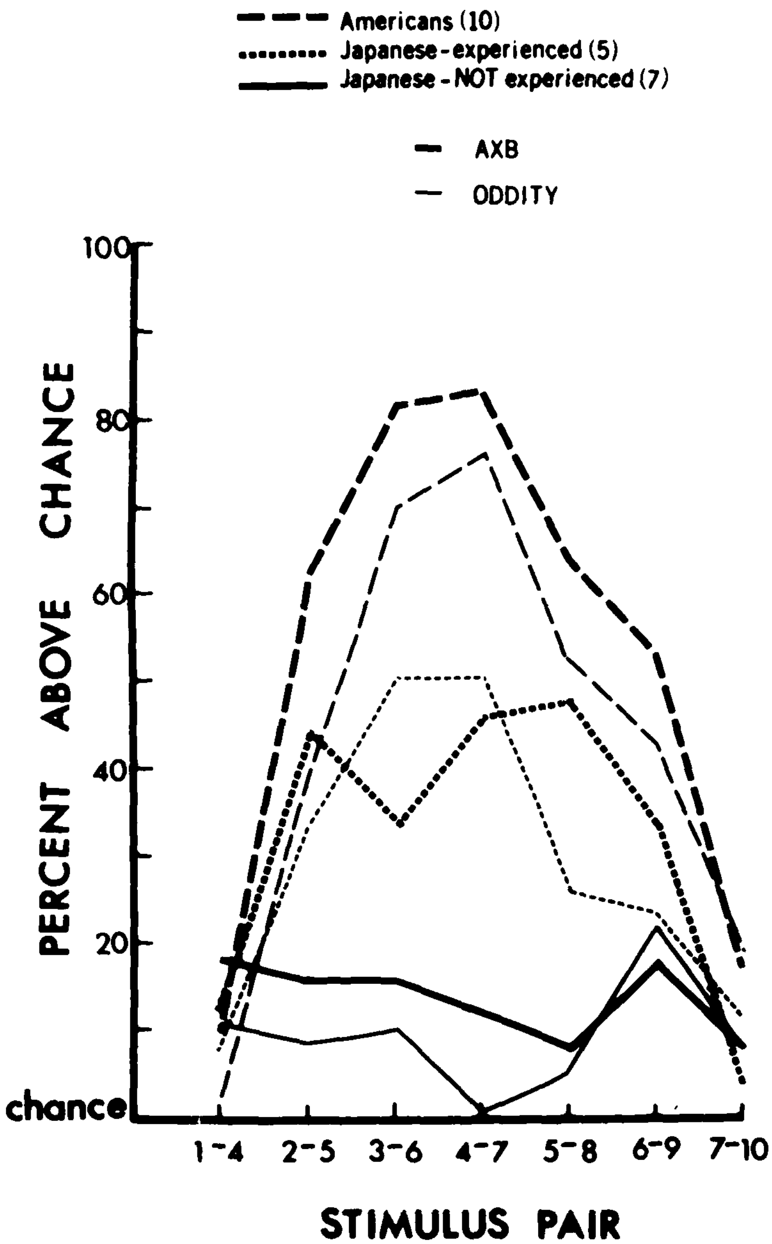


Figure 8. Comparison of the AXB and oddity tests on above-chance performance by the three groups of subjects.

The sum of the results from the comparison of discrimination tasks does not lend support to the notion that the reason that AXB performance exceeds oddity performance is because the former task allows subjects better access to pre-categorical (nonphonetic, or auditory) information, at least not when judgments are being made on stimuli whose characteristics approach the acoustic properties found in natural speech. If the auditory memory/phonetic bias picture were correct, all three groups should have fared better on AXB than on oddity judgments. Also, a significant Test x Stimulus Pairs interaction should have been found for the Americans and the Experienced Japanese, indicating that within-category judgments were improved on the AXB task relative to between-category judgments. Furthermore, the NOT Experienced Japanese should have shown even greater task effects than the other two groups, since they could use nonphonetic auditory memory but could not rely on phonetic categorization. Instead of the predictions being supported, the pattern of AXB-oddity comparisons across the three groups suggests that performance on both tests reflects the effects of phonetic perception. The only group that showed significantly higher AXB than oddity performance was the group that was most experienced with /r/ and /l/ as a phonemic contrast--the Americans. Recall also that this was the only group whose obtained performance on both discrimination tests was better than predicted by their identification data, and that they only showed better than predicted performance on within-category comparisons that differed in "category goodness." In addition, the nonsignificant Test x Stimulus Pairs interaction for the Americans indicates that their AXB advantage was not due to better accessing of nonphonetic auditory information, but rather that it derived from some improvement in access to specifically phonetic information.

DISCUSSION

This study investigated categorical perception of /r/ and /l/ by native Japanese speakers residing in the U.S. who had had varying amounts of experience in English conversation with native speakers. Bilingual Japanese speakers who were not experienced in English conversation with natives showed, with one exception, near-chance performance across the /r/-/l/ series in the identification task and correspondingly low performance on the discrimination tests. These results corroborate and, for the oddity discrimination test, replicate the earlier Miyawaki et al. (1975) findings with a new stimulus series that provided redundant cues for the phonetic contrast.

The group of focal interest, those bilingual Japanese speakers with relatively intensive conversational experience, performed more similarly to the American-English controls than to their less experienced Japanese counterparts. The identification function for each of the Experienced Japanese showed a sharp category boundary that was nearly indistinguishable from those of the American-English controls (see Figure 6). Discrimination results were well predicted from the identification data, showing significant peaks in performance at the category boundary for both tests. These results are most encouraging, for they demonstrate that native Japanese speakers learning to converse in English as adults can achieve phonetic categorization of /r/ and /l/ that approximates the categorization behavior of native English speakers.

It is appropriate at this point to discuss the unusually excellent performance of one nonexperienced Japanese subject, M. K. Much to our surprise, his performance on the three tests was more similar to the Experienced Japanese group than to the NOT Experienced group (see Figure 7). The major distinction between his data and those of the Experienced Japanese was that the form of his discrimination functions were not predicted by his /r/-/l/ identification results, suggesting that his ability to discriminate the stimuli may not have been directly tied to his phonetic classification of them. However, an alternative explanation to his uncorrelated discrimination responses has not been ruled out. During the identification test, only one stimulus is presented and a categorization response is noted immediately. In contrast, the discrimination tasks require that two or three sounds be held in memory over several seconds before discrimination judgments are made. Under these memory demands, unstable phonetic representations for these sounds might be disrupted easily, resulting in less consistent performance. We suspect that M. K.'s consistent identification of /r/ and /l/ shows an unusual sensitivity to phonetic distinctions; however, without additional measures of his perceptual behavior, his performance remains an interesting anomaly.

This study has demonstrated that some native Japanese speakers learning English as adults are capable of categorically perceiving /r/ and /l/ in a manner similar to native English speakers. Differences in performance between the Experienced and NOT Experienced groups were correlated with differences in conversational experience; however, we cannot rule out a host of variables (e.g., motivation to learn) that might account for differences in performance between the two Japanese groups. Because this study was not designed to test the longitudinal effects of experience, with pre- and post-testing of the same subject on perception of /r/ and /l/, we must infer that the Experienced group represented typical native Japanese speakers, and that they at one time failed to perceive /r/ and /l/ categorically. We are fairly confident that this is the case since each subject was asked about previous problems with /r/ and /l/, and they all reported having great difficulty with this contrast initially, as well as reporting a gradual improvement over time.

The design used here cannot directly answer questions about whether and what kinds of experience produce the change toward categorical perception of phonetic contrasts. Laboratory training studies have had some success in improving /r/-/l/ perception by native speakers of Oriental languages in which the contrast is not phonological. For example, Gillette (Note 4) reports significant improvement in natural /r/ and /l/ identifications by Japanese and Korean native speakers following several weeks of intensive training with natural speech. Dittmann and Strange (Note 5) have used a same-different discrimination task with feedback, and produced a change in perception of a synthetic /r/-/l/ series from uniformly poor discrimination to categorical perception by native Japanese speakers.

Future research should be directed toward discovering the perceptual strategies speakers use in their acquisition of this contrast, and determining the conditions that best facilitate acquisition of this contrast by second language learners. Some laboratory training studies currently employ repetitions of minimal pairs of words, natural or synthetic, in listening tasks that require subjects to perform highly differentiated analyses at the level of distinctive features. In accordance with results from first language learners

(cf. Menyuk & Menn, 1979), it may be more efficacious, in initial learning of a non-native contrast by adults, to approximate the first language learning situation in which words are presented in natural speech in sentence contexts and related to objects and events, thus maximizing information at a number of linguistic levels. Following experience with /r/ and /l/ under these conditions, redundant information could be reduced systematically until subjects are required to perform under the most demanding situation, that of making a perceptual distinction between minimal pairs.

REFERENCE NOTES

1. Mochizuki, M. The identity of /r/ and /l/ in natural and synthesized speech. Unpublished manuscript, University of Pennsylvania, 1980.
2. Miyawaki, K. A study of lingual articulation by use of dynamic palatography. Unpublished masters thesis, University of Tokyo, 1973.
3. Price, P. J. A cross-linguistic comparison of flaps in Japanese and in American English. Unpublished doctoral dissertation, University of Pennsylvania, 1981.
4. Gillette, S. Contextual variation in the perception of L and R by Japanese and Korean speakers. Minnesota Papers in Linguistics and the Philosophy of Language, 1980, 6, 59-72.
5. Dittmann, S., & Strange, W. Effects of discrimination training on the perception of /r/ and /l/ by a native Japanese speaker. Paper presented at the 7th annual Minnesota Regional Conference on Language and Linguistics, May 8-9, 1981.

REFERENCES

Abramson, A. S., & Lisker, L. Discriminability along the voicing continuum: Cross language tests. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, 1970.

Abramson, A., & Lisker, L. Voice timing perception in Spanish word-initial stops. Journal of Phonetics, 1973, 1, 1-8.

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 1981, 29, 191-211.

Dalston, R. M. Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. Journal of the Acoustical Society of America, 1975, 57, 462-469.

Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Lisker, L., & Abramson, A. S. A cross-linguistic study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.

Lisker, L., & Abramson, A. S. Some effects of context on voice onset time in English stops. Language and Speech, 1967, 10, 1-28.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, 1970.

Menyuk, P., & Menn, L. Early strategies for the perception and production of

words and sounds. In P. Fletcher & M. Garman (Eds.), Language acquisition. Cambridge: Cambridge University Press, 1979, 49-70.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 1975, 18, 331-340.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Acoustic cues for the perception of initial /w,r,l/ in English. Word, 1957, 13, 25-43.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. Cross language study of vowel perception. Language and Speech, 1969, 12, 1-23.

Terbeek, D. A cross-language multi-dimensional scaling study of vowel perception. Working Papers in Phonetics (University of California at Los Angeles), 1977, 37.

Williams, L. The voicing contrast in Spanish. Journal of Phonetics, 1977, 5, 169-184.

FOOTNOTES

[1]The steps were not exactly equal because of the hardware limitations of the OVE-IIIc synthesizer. In all cases, the deviations from exact equality in step sizes were only a few Hz.

Appendix A. Nominal parameter values for the /rɑk/-/lɑk/ stimulus series.

Numbers represent the duration (in milliseconds) of the initial steady state (SS) and the transition (Tran) of the first formant (F1), the center frequencies of the second (F2) and third (F3) formants at the beginning of the syllables (start), and the center frequency of F3 at the point of inflection 35 ms into the syllable (T = 35).

Stimulus number	F1 Duration (ms) SS	F1 Duration (ms) Tran	Formant Center Frequencies (Hz) F2 Start	Formant Center Frequencies (Hz) F3 Start	Formant Center Frequencies (Hz) F3 (T=35)
1	14	49	1067	1477	1576
2	14	49	1083	1611	1694
3	21	42	1099	1731	1808
4	21	42	1115	1847	1915
5	28	35	1131	1972	2029
6	28	35	1147	2104	2135
7	35	28	1156	2229	2262
8	35	28	1172	2345	2362
9	42	21	1189	2466	2484
10	42	21	1207	2594	2594

Constant Portion of Stimuli

Formant Center Frequencies (in Hz)

F1 Start	Vowel F1	Vowel F2	Vowel F3	Final Closure F1	Final Closure F2	Final Closure F3
349	621-707	1198-1233	2557	621	1288	2104

INFLUENCE OF VOCALIC CONTEXT ON PERCEPTION OF THE [ʃ]-[s] DISTINCTION: V. TWO WAYS OF AVOIDING IT

Bruno H. Repp

Abstract. Three experiments investigated the conditions under which fricative perception is influenced by following vocalic context. In Experiment 1, a reaction-time task, listeners showed no such influences, suggesting that they reached decisions about the fricative category before processing the vocalic context. In Experiment 2, a fixed-standard AX discrimination task employing synthetic fricative noises from a [ʃ]-[s] continuum, listeners successfully discriminated fricative noises in isolation but shifted to a phonetic (categorical) mode of perception when vocalic context was added. Their response patterns changed systematically with the nature of the context. In Experiment 3, the subjects listened first to pairs of isolated noises immediately followed by the same noises in context. When, subsequently, only noises in context were presented for discrimination, most of the subjects performed noncategorically and were no longer influenced by different vocalic contexts. These experiments demonstrate the availability of different perceptual strategies in listening to speech.

In a recent study (Repp, 1980a), I used synthetic noises from a [ʃ]-[s] continuum, followed by vocalic portions known to influence the location of the [ʃ]-[s] boundary in an identification test. The stimuli were presented in AXB and fixed-standard AX discrimination tasks. The majority of naive subjects perceived these fricative-vowel syllables fairly categorically in both tasks; that is, discrimination functions followed the patterns predicted from identification scores and showed shifts contingent on the nature of the vocalic portion. However, two subjects achieved much better discrimination scores than the rest, and so did three experienced listeners who participated in the AX task. These listeners, who (judging from their higher accuracy, pattern of responses, and subjective reports) successfully followed the nonphonetic strategy of restricting attention to the spectral properties of the fricative noise, were not influenced by different vocalic contexts. These results supported the hypothesis that influences of vocalic context on fricative identification are tied to a phonetic mode of perception.

EXPERIMENT 1

The experiment just summarized suggests that, when listening to fricative-vowel syllables in a phonetic mode, subjects process the vocalic portion

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

before making a decision about the fricative category. However, this observation may have only limited generality. On one hand, the fricative noises used were highly ambiguous, and the resulting uncertainty may have delayed the phonetic decision, thus permitting it to be influenced by the following context; on the other hand, the discrimination tasks did not demand rapid phonetic decisions. It was the purpose of the present Experiment 1 to investigate whether vocalic context effects would be obtained in a reaction-time task with unambiguous fricative noises. It is known that, in a standard identification task, natural [s] and [ʃ] noises are fairly immune to contextual effects, i.e., they are generally sufficient cues for accurate identification of the fricative consonant (Harris, 1958; LaRiviere, Winitz, & Herriman, 1975). However, if listeners follow a strategy of waiting for the end of the fricative noise before making a decision, context effects might be revealed in an analysis of response latencies.

As a further test of whether listeners wait for the vocalic stimulus portion before making a decision about fricative identity, the duration of the fricative noise portion was varied. The hypothesis that listeners do wait would be supported if an increase in noise duration led to an equivalent increase in reaction time, regardless of whether or not vocalic context has any effect (cf. Repp, 1980b, for a similar design).

Method

Subjects. Ten paid student volunteers participated. Some of them had been subjects in earlier, similar experiments, while others were relatively naive.

Stimuli. The utterances [sɑ], [ʃɑ], [su], [ʃu] were recorded by a female speaker (FBB). Three different tokens of each syllable were selected, digitized at 20 kHz, and low-pass filtered at 9.8 kHz. Within each of the three sets of four syllables, the aperiodic and periodic stimulus portions were separated and recombined in all possible ways, leading to three sets of 16 stimuli, 48 in all. A second set of 48 stimuli was obtained by shortening each fricative noise by 50 msec.[1] The 96 stimuli were recorded on tape in three randomized sequences with interstimulus intervals of 1.5 sec. Each sequence was preceded by four warm-up stimuli that were ignored in data analysis.

Procedure. Subjects were seated in a sound-insulated booth and rested their index fingers on two telegraph keys labeled "s" and "sh". They listened over Telephonics TDH-39 earphones and were instructed to identify the fricative consonants as quickly as possible by pressing one of the keys. The hand-response assignment was counterbalanced between subjects. The stimulus tape was played back twice on a Crown 800 tape recorder located in an adjacent room; thus, the subjects listened to 6 blocks of 100 stimuli, each lasting about 3 minutes. Subjects were permitted to stop the tape by remote control and take a rest between blocks, if desired. Reaction times were measured by a Hewlett-Packard 5302A 50MHz universal counter and printed out on a Hewlett-Packard 5150A thermal printer. The timer was triggered by a tone recorded on the other tape channel and synchronized with fricative noise onset.

Analysis. The first block served as practice; only the data from blocks 2-6 were considered. Each subject gave 5 responses to each of the 3 tokens of each of 32 stimuli. Medians of the 5 response times were calculated (excluding errors) before computing means across tokens. These means were analyzed in a 5-way ANOVA with the factors: (A) fricative noise duration, (B) fricative category, (C) vowel category, (D) original fricative (i.e., the category of the fricative that originally preceded the periodic stimulus portion), and (E) original vowel (i.e., the category of the vowel that originally followed the fricative noise).

Results and Discussion

As expected, errors were rare; they ranged from 0.6 to 6.3 percent across subjects. Thus, the aperiodic stimulus portions provided sufficient information for fricative identification. The average reaction times of individual subjects ranged from 334 to 590 msec; the grand mean was 449 msec.

If vocalic context had any effect, decision times should have been slowed down when a fricative noise was followed by either a vowel from a different original fricative context, i.e., by a vowel containing formant transitions appropriate for the other fricative category (reflected in the BxD interaction of the ANOVA) or by a vowel from a category different from that of the original vocalic context of the fricative noise (CxE interaction). However, neither interaction was significant, $F(1,9) < 1$.

The effect of fricative noise duration reached significance, $F(1,9) = 7.6$, $p < .05$. Long fricative noises took longer to respond to than short noises, but the average difference was only 8 msec, instead of the expected 50 msec. This suggests that the listeners did not wait for the vocalic portion before making a decision.

The only highly reliable effect was a main effect of factor E (original vowel), $F(1,9) = 44.6$, $p < .001$: Noises from original [ɑ] context were responded to faster (by 14 msec) than noises from original [u] context. There was a durational difference between noises from the two context: On the average, noises from [ɑ] context were 34 msec longer than noises from [u] context. Again, however, there is a mismatch in the magnitudes of the two temporal differences, suggesting that the effect of original vowel was not an effect of noise duration. Perhaps, fricative noises from [u] context were perceived as less typical of their respective categories because their spectrum was lowered by anticipatory lip rounding.

Another way of looking at effects of fricative noise duration, which was not confounded with any experimental manipulations, was to examine differences in reaction time to the three individual tokens of the fricative noises from the four original utterances. Combining all contexts in which a given noise occurred, as well as its long and short versions, between-token differences were tested for significance in four separate analyses of variance. The token effect was significant only for noises deriving from [sɑ], $F(2,18) = 6.8$, $p < .01$. This was also the only case of a monotonic and positive relation between noise duration and reaction time; but, once again, the latency difference between the two extreme noises (33 msec) was smaller than the difference in noise duration (56 msec).

Thus, this study does not support the hypothesis that listeners wait for the onset of the vocalic portion; on the contrary, they apparently based their decisions on the fricative noise alone and ignored the vocalic context. There are two possible explanations. One is that the subjects adopted an auditory rather than a phonetic criterion and based their decisions on the pitch quality of the noise, which does not seem to be affected by vocalic context (Repp, 1980a, and Exp. 3 below). The other possibility is that the subjects were in a phonetic mode but accumulated information right from the beginning of each stimulus and initiated a decision as soon as this information was sufficient, which occurred some time before the vocalic portion came on. The second explanation is more plausible on the following grounds. First, the task demanded identification of the fricative consonants as "s" or "sh", which furthered a phonetic mode of perception. Second, all ten subjects of the present study also participated in Experiment 2, described below, which required discrimination of fricative noises in context, and all subjects perceived these stimuli categorically, i.e., they were not able to pay attention to the spectral qualities of the noise and to ignore the vocalic context. Third, Whalen (Note 1) recently demonstrated that effects of vocalic context on fricative identification latencies do emerge when identification of the fricative <u>and</u> of the following vowel is required in a four-choice task, i.e., when listeners are forced to wait for the vocalic portion.

Thus, the tentative conclusion from Experiment 1 is that listeners accumulate phonetic information continuously, and if the task requires that decisions be made at the phonetic level, such decisions can be initiated as soon as sufficient information has been collected (cf. Repp, 1980b).[2] Presumably, every listener possesses this ability, which is distinct from the ability to gain access to <u>auditory</u> properties of a signal portion such as the pitch of the fricative noise. My earlier experiments (Repp, 1980a) showed that this latter ability is not immediately present in most listeners. The following two studies examined what sort of training might enable listeners to acquire it.

EXPERIMENT 2

In Experiments 2 and 3, I attempted to teach a group of naive subjects to discriminate fricative noises in context, i.e., to abandon the phonetic (categorical) mode of perception in favor of an auditory (noncategorical) strategy. Because of the relative accessibility of the auditory differences involved, it was expected that little training would be necessary to transform categorical listeners into noncategorical listeners. In fact, the ability to focus attention on the noise portion of fricative-vowel stimuli might be discovered rather than slowly learned, as suggested by the extremely accurate performance of two naive listeners in my earlier study (Repp, 1980a). The first study examined whether it would be sufficient for subjects to hear and discriminate the fricative noise stimuli in isolation.

Method

<u>Subjects</u>. The same ten subjects as in Experiment 1 participated.

<u>Stimuli</u>. The stimulus tapes were the same as in Experiment 2 of Repp (1980a), and the reader is referred to that earlier report for details. The

stimuli were synthetic noises from a 7-member [ʃ]-[s] continuum, followed by one of two natural-speech periodic portions, [(ʃ)ɑ] or [(s)u]. The first of these (an [ɑ] with formant transitions appropriate for [ʃ]) biased fricative identification towards "sh", whereas the second (an [u] with formant transitions appropriate for [s]) biased fricative identification towards "s". The stimuli were presented in a fixed-standard AX format. Stimulus 4 on the noise continuum served as the standard. In each stimulus pair, it was followed by a comparison stimulus which could be any of the seven stimuli, with equal probability. There were four different conditions. In two conditions, the standard and the comparison always had the same periodic portion--[(ʃ)ɑ] in one condition and [(s)u] in the other. In the other two conditions, the periodic portions were always different--[(ʃ)ɑ] for the standard and [(s)u] for the comparison in one condition, and the reverse assignment in the other. Each condition contained 24 repetitions of the 7 possible stimulus pairs, of which the first 4 served as practice and were not scored. In addition to these four tapes, a tape containing isolated noise stimuli in the same fixed-standard AX format was prepared.

Procedure. All subjects listened first to the two conditions (order counterbalanced across subjects) in which standard and comparison noise stimuli were followed by the same periodic portions. Subsequently, they listened to the isolated noises. Finally, the two conditions in which different periodic portions followed the standard and comparison noises were presented (order counterbalanced across subjects), to test whether anything had been learned from discriminating the noises in isolation. All tapes were presented in a single session, and the responses were "s" and "d". The subjects were fully informed about the nature of the stimuli and were instructed to pay attention to differences in the noise portion only and to ignore the vowel.

Results and Discussion

The results are displayed in Figure 1. In the left-hand panel, the functions for the first two conditions replicate the pattern found for the categorical listeners in Experiment 2 of Repp (1980a). In fact, all ten subjects fit that pattern; there were no noncategorical listeners in the present study. In the right-hand panel of Figure 1, we see that the subjects did rather well with the isolated noises; clearly, these stimuli were not categorically perceived. Despite this success, however, all subjects apparently reverted to a phonetic mode of perception in the remaining two conditions, whose pattern of results again resembles that found in Experiment 2 of Repp (1980a) for categorical listeners.

The statistical analysis of the four vocalic-context conditions confirmed that, as in Experiment 2 of Repp (1980a), the periodic portion of the standard stimulus had a significant effect on the shape of the discrimination function, $F(6,54) = 4.7$, $p < .001$, and that there were more "different" responses to pairs of stimuli differing in their periodic portions than to pairs that had the periodic portion in common, $F(1,9) = 13.0$, $p < .01$. The latter effect was confounded with practice and may reflect some slight improvement in the course of the experiment, in addition to a response bias induced by the relationship between the irrelevant stimulus portions. Clearly, however, the subjects did not become noncategorical listeners.

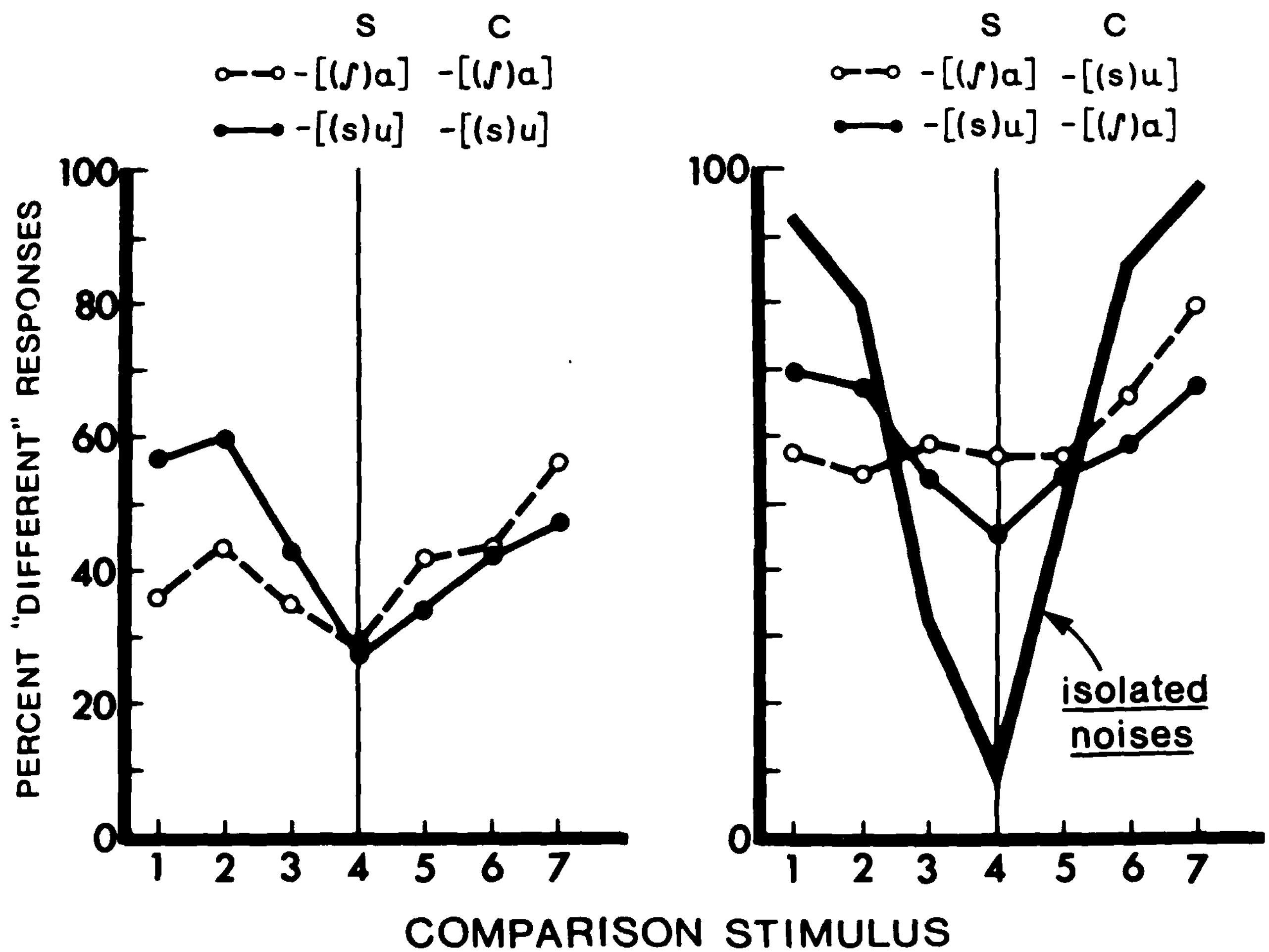


Figure 1. Fixed-standard AX discrimination before (left-hand panel) and after (right-hand panel) discrimination of isolated noises (Exp. 2).

The pattern of the data is to be interpreted as follows: When both noises in a pair were followed by [(ʃ)ɑ], the standard stimulus was categorized as "sh"; consequently, it was difficult to discriminate from more [ʃ]-like noises (stimuli 1-3), but discrimination from more [s]-like noises (stimuli 5-7) improved with their physical distance from the standard because they crossed the phonetic category boundary. Conversely, when both noises were followed by [(s)u], the standard was categorized as "s"; consequently, discrimination from more [s]-like noises (stimuli 5-7) was poor, but discrimination from more [ʃ]-like noises (stimuli 1-3) improved with their physical distance from the standard because they crossed the phonetic category boundary (cf. left-hand panel of Fig. 1). In the two conditions where standard and comparison noises were followed by different periodic portions (right-hand panel of Fig. 1), the situation is similar, but the minimum percentage of "different" responses might be expected to shift away from the center (stimulus 4): A comparison stimulus followed by [s(u)] must be more [ʃ]-like (and one followed by [(ʃ)ɑ] must be more [s]-like) than the standard to sound most similar to it. It is interesting to note that this latter effect was absent: Only the periodic portion of the standard, but not that of the comparison stimulus, had any influence on listeners' responses. (This was also true in the earlier data of Repp, 1980a.) This finding is reminiscent of the absence of vocalic-context effects in Experiment 1: Subjects may have been able to initiate the phonetic decision and comparison before processing the periodic portion of the comparison stimulus, but they could not ignore the periodic portion of the standard stimulus, which had to be held in memory until the comparison stimulus arrived.

EXPERIMENT 3

The subjects in Experiment 2 were not able to transfer the discriminatory skill exhibited with isolated noises to the same noises in vocalic context. This suggests that a better awareness of the auditory dimension on which the noises differ is not sufficient to accomplish the task. What may be required, in addition, is the ability to segregate the noise from the following periodic portion and thereby to escape from the phonetic mode of perception. The present study tried out one of several possible methods that might teach listeners this skill.

Method

Seven of the ten subjects in Experiment 2 returned for this experiment. In addition, two new volunteers participated. All subjects listened first to a training tape. On this tape, two of the previous conditions were interleaved: On each trial, a pair of isolated noises was followed, after 2 sec, by a pair of exactly the same noises in the [(s)u] context. The subjects were instructed to listen to the isolated noises, to determine the nature of the difference (if any), and then to verify for themselves that exactly the same difference existed between the noises in the syllables. During the first block of 28 trials, the subjects looked at an answer sheet that specified exactly which noise stimuli occurred on each trial. (The nature and arrangement of stimuli was first explained in detail.) During the remaining five blocks, the subjects responded "s" or "d" after listening to both pairs on each trial. They were urged to continue to compare the noise differences in the two pairs.

Following this training condition, the subjects listened to three of the tapes used earlier, in a fixed order. On the first tape, both noise stimuli were followed by [(s)u]; thus, this condition was identical with the training condition, except that pairs of syllables were no longer preceded by pairs of isolated noises. Next, subjects listened to the condition in which the standard was followed by [(s)u] and the comparison stimulus by [(ʃ)a], and finally to the condition in which the standard was followed by [(ʃ)a] and the comparison stimulus by [(s)u].

Results and Discussion

Preliminary inspection of the results indicated that the new training method was quite successful, but three subjects seemed to benefit much less than the other six (who included the two newcomers). Therefore, the results are displayed separately for "poor" and "good" subjects in Figure 2.

It is evident that the three poor subjects had some trouble in the training task; in particular, they missed out on identical pairs, producing almost 50 percent false alarms (i.e., incorrect "different" responses). Their performance in the three subsequent tests was extremely poor, due to the even higher false-alarm rates (over 70 percent). However, there were no clear effects of vocalic context. The high false-alarm rates were almost certainly due to the subjects' knowledge (from the training task and from the preceding instructions) of the true proportion of "same" trials (viz., only one out of seven). However, they also indicate that these subjects found it more difficult than the others to discriminate isolated noises.[3] For this reason, they benefited less from the training task, which served its purpose only to the extent that the differences between isolated noises could be detected. On the other hand, the apparent absence of vocalic context effects suggests that, rather than persisting in a phonetic mode, these subjects perhaps did learn to segregate the noise portions from their vocalic contexts but then could not easily detect the spectral differences between them (or, rather, their spectral identity). In other words, the epithet "poor," rather than "categorical," seems to be appropriate for these subjects in this experiment.

The six good subjects, on the other hand, were obviously very accurate in the training task and benefited from that experience. Although their false-alarm rates in the vocalic-context conditions were higher than those of the noncategorical subjects in Experiment 2 of Repp (1980a) (presumably because the present subjects knew about the infrequent occurrence of "same" trials), the discrimination functions were V-shaped and clearly different from those of the categorical subjects in previous experiments. (Note that four of the six good subjects had participated and produced categorical results in Exp. 2.) In fact, when the average scores were converted into d' values, they were slightly higher than those of the noncategorical subjects in Experiment 2 of Repp (1980a) (who included the author and two other investigators), indicating remarkable success in the task. There was no clear effect of vocalic context. This was confirmed in an analysis of variance of the scores in the two conditions with unequal periodic portions (lower right-hand panel of Fig. 2), $F(6,30) = 2.1$, $p > .05$. There was no indication here of any reversed context effect, as in Experiment 2 of Repp (1980a), although three of the six subjects showed a tendency in that direction. There was no significant effect of vocalic context in a combined analysis of all nine subjects in the present experiment, $F(6,48) = 1.5$.[4]

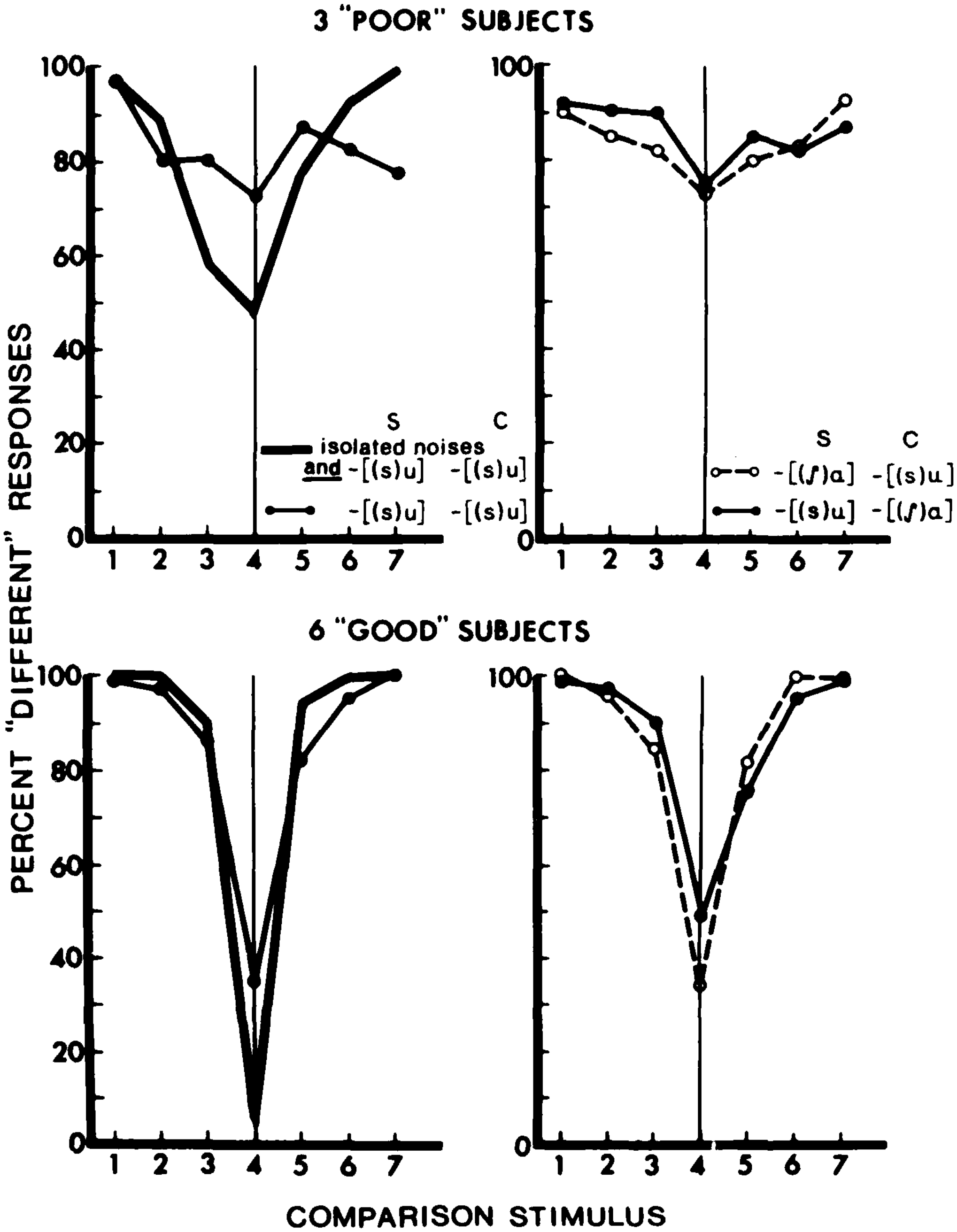


Figure 2. Fixed-standard AX discrimination performance of three "poor" and six "good" subjects after listening to a training tape (Exp. 3).

GENERAL DISCUSSION

The training task in Experiment 3 was effective and sufficient to convert most listeners from a categorical (phonetic) to a noncategorical (nonphonetic or auditory) mode of listening. The fact that the majority of subjects became as accurate as experienced listeners after only 25 minutes of self-training is consistent with the suggestion, made above, that the skill of segregating and discriminating fricative noises in vocalic context is discovered rather than slowly learned; that is, it reflects a qualitative change in perceptual processing. Quantitative improvements, such as might occur with further practice, are contingent on that change.

In support of this claim, it should be noted that the distribution of accuracy levels across subjects was quite bimodal. Listeners were either very accurate or very poor in discriminating fricative noises in context; there was not a single subject who performed at an intermediate level. (Such a level would be expected only if a listener alternated between the two strategies.) Also, one of the categorical listeners in Experiment 2 apparently switched strategies ("caught on") between the last two conditions, which resulted in a sudden and dramatic improvement in performance.

The present data provide further support for the hypothesis that effects of vocalic context on fricative identification are tied to a phonetic mode of perception. They suggest strongly that there are two different strategies of listening to fricative-vowel syllables, one auditory (noncategorical) and the other phonetic (categorical). Regular vocalic context effects occur only in the phonetic mode; however, they may not be manifest when stimuli and task permit subjects to make a rapid phonetic decision before processing the context (Exp. 1). Contextual effects reflect implicit knowledge of articulation and coarticulation and/or their acoustic consequences. To bring this knowledge to bear on some auditory input is tantamount to being in a phonetic mode of perception: We perceive speech in terms of what our brain knows about it. Similarly, we perceive nonlinguistic auditory attributes of speech with reference to what we know about nonspeech sounds. The frame of reference adopted for a particular stimulus is a joint function of stimulus structure and listener strategy.

The phonetic and auditory modes are available, in principle, for any speechlike stimulus. They may even be used simultaneously. However, since the phonetic mode is the natural way of dealing with speech, and since the auditory properties of speech are often unfamiliar and require the listener to pay attention to fine detail, special laboratory tasks may be necessary to elicit the auditory listening strategy. Fricative-vowel syllables differ from, say, stop-consonant-vowel syllables in that some of their auditory properties (e.g., the pitch of the fricative noise) are easier to access and discriminate (as compared to, e.g., the "pitch" of formant onsets or the duration of aspiration). The relative accessibility of an auditory property is largely governed by stimulus factors: Auditory judgments of the pitch of fricative noises are easiest to make when the noises occur in isolation, more difficult in fricative-vowel syllables, and probably even more difficult when the fricatives occur intervocalically or are embedded in fluent speech. (The fact that the fric.tive noises in the present studies were synthetic may also have been a facilitating circumstance.) Task factors, such as interstimulus

intervals and stimulus uncertainty, naturally play a role, too. In principle, any auditory property of speech can be detected and discriminated within the limits set by the auditory system, but listeners may have to learn how to gain access to the relevant property. They may have to reorganize their percept in the process (e.g., "segregate" the noise portion from the following vocalic portion), which involves perceptual skills that *need to be acquired* or elicited by appropriate instructions.

REFERENCE NOTE

1. Whalen, D. H. Effects of non-essential cues on the perception of English [s] and [š]. Paper presented at the 101st Meeting of the Acoustical Society of America, Ottawa, Ontario, May 1981.

REFERENCES

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1958, 1, 1-7.

LaRiviere, C., Winitz, H., & Herriman, E. The distribution of perceptual cues in English prevocalic fricatives. Journal of Speech and Hearing Research, 1975, 18, 613-622.

Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction: IV. Two strategies in fricative discrimination. Haskins Laboratories Status Report on Speech Research, 1980, SR-63/64, 121-138. (a)

Repp, B. H. Accessing phonetic information during perceptual integration of temporally distributed cues. Journal of Phonetics, 1980, 8, 185-194. (b)

FOOTNOTES

[1]These 50-msec segments were removed from the center of each noise portion, so as not to interfere with its onset or offset. The original noise durations ranged from 170 to 244 msec, the reduced durations from 120 to 194 msec. The shortest noises had a somewhat affricate-like quality, but it is unlikely that this influenced reaction times.

[2]The data leave open the possibility that subjects waited for the fricative noise to end (but not for the periodic portion to begin) before initiating a decision. The increase in latencies occasioned by a long noise may have been partially offset by a reduction in uncertainty due to the larger amount of information carried by a longer noise. The resulting faster decisions may have attenuated the manifest effects of noise duration. However, this possibility remains rather implausible.

[3]These three subjects did not do very well either in the isolated-noise condition of Experiment 2. This seems to rule out the alternative explanation that they tended to base their judgments on the syllables rather than on the isolated noises in the present training condition.

[4]There was probably a general effect of vocalic context: Performance was more accurate with isolated noises than with noises in any vocalic context. This difference may be ascribed in part to interference between stimulus

portions in auditory memory. However, performance in the training condition was also favored by shorter noise-to-noise intervals in pairs of isolated noises (the interstimulus interval was 2 sec in all conditions) and by the opportunity to extract additional information from the following pair of syllables.

GRAMMATICAL PRIMING OF INFLECTED NOUNS

G. Lukatela,+ A. Kostić++ and M. T. Turvey+++

Abstract. In normal linguistic usage, the inflected nouns of Serbo-Croatian are usually preceded by prepositions that help to specify which particular grammatical case is intended and to stress the noun's function in the sentence. In a lexical decision task it was demonstrated that lexical decision times to nouns in a grammatical case that demands a preposition were faster when the preposition was appropriate to the case than when it was either inappropriate to the case or a nonsense syllable. This result lends support to the intuition that priming can occur among sentential components.

It is easily demonstrated that naming a word is facilitated by the prior occurrence of the word itself or a semantically related word (for example, Fischler, 1977; Meyer, Schvaneveldt, & Ruddy, 1975; Scarborough, Cortese, & Scarborough, 1977), but it is debatable whether such facilitation occurs in normal linguistic usage. Semantic priming of lexical items is most commonly demonstrated in the context of word lists, and in the view of Forster (1976) it is a phenomenon that may well be restricted to this context. Forster sees related words as interconnected or cross-referenced in the lexicon and this cross-referencing is the basis for semantic facilitation effects. Given this view, Forster (1976) is dubious that sentence fragments can provide the semantic context that primes lexical entries; rarely are individual words in sentences of English semantically related. Forster reports that words that were predictable from a sentence context were not named faster than words that were less predictable. But there are some strong hints to the contrary (e.g., Blank & Foss, 1978; Morton & Long, 1976; Schuberth & Eimas, 1977; Underwood, 1977).

A procedure that has proved extremely sensitive to short-term facilitatory—and inhibitory (see Neely, 1977)--effects of one linguistic item on another is the lexical decision task. Quite simply, in this task a subject is shown a string of letters and is required to respond as quickly as possible to its lexical status; that is, the subject decides whether the letter string is a word. The lexical decision task is used in the experiment reported here, which looks at the possibility of facilitating the processing of inflected nouns through the prior presentation of an appropriate preposition.

+University of Belgrade.
++Also University of Belgrade and University of Connecticut.
+++Also University of Connecticut.
Acknowledgment. This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade and in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

Inflection is the major grammatical device of Serbo-Croatian, Yugoslavia's principal language. A noun 'system' in Serbo-Croatian consists of seven cases, both in the singular and in the plural. Excluding the nominative and vocative cases, each grammatical case has a number of possible meanings. The particular meaning is specified either by a preposition or by the sentence context. The grammatical cases of a Serbo-Croatian noun are formed by adding to the root form an inflectional morpheme, namely, a suffix consisting of one syllable of the vowel or vowel-consonant type. Inflecting the noun may also involve deleting a vowel and palatalizing a consonant. At all events, in normal linguistic usage the grammatical cases formed are preceded by a preposition that serves (1) to specify which particular grammatical case is intended (where more than one grammatical case is represented by a given orthographic and phonological structure) and (2) to specify which particular meaning of the grammatical case is intended (where more than one meaning is associated with a given grammatical case). In other words the relationship of a preposition to a grammatical case is one of complementation. In isolation the grammatical information revealed by a particular case (with the exception of the nominative and vocative) is equivocal. This equivocality is reduced through a preposition that specifies the case and clarifies its role in the sentence, pointing to the particular meaning it is to assume. And it is reduced further by the overall context of the sentence.

Significantly, the preposition/inflected noun relation is more properly described as a grammatical or functional relation rather than as a semantic association. We would not, in short, expect prepositions and inflected nouns to be cross-referenced in the lexicon in the same manner that Forster (1976) conceives semantic relatives to be cross-referenced. Indeed, there is some reason to believe that for English the internal representation of function words (prepositions and the like) is not common with the internal representation of content words. Thus, phonemic dyslexics who are generally unable to read pseudowords are generally successful at reading words, with the curious exception of function words. Apparently, phonemic dyslexics relate to function words as if they were, like pseudowords, without representation in the lexicon and, therefore (given the inability to derive phonology rulefully from script) unreadable (Patterson & Marcel, 1977). In a related observation Bradley (1978) notes that whereas lexical decision on content words is faster the higher the frequency of the word, lexical decision on function words is independent of frequency of occurrence.

The preposition/inflected noun relationship is significant in another way. As noted, the inflected nouns of Serbo-Croatian are most usually preceded in normal spoken and written discourse by an appropriate preposition. A preposition, therefore, is quite legitimately a "sentence fragment," and if a facilitation of the lexicon by prepositional primes can be demonstrated, then it is reasonable to assume that in the more natural setting of sentence perception (as contrasted with word-list perception), parts of a sentence perceptually facilitate other parts. There is already good reason to believe that the preposition/inflected noun relation is significant in auditory sentence processing by reducing the reliance on preserving or attending to word order. In Serbo-Croatian, prepositions and inflected endings serve as local markers of a word's role and appear to contribute to the more rapid acquisition of sentence processing strategies by young listeners of Serbo-Croatian as compared to young listeners of English (Ammon & Slobin, 1979).

We chose to investigate the effect of appropriate, inappropriate, and nonsense prepositions on lexical decision to Serbo-Croatian nouns in three grammatical cases--the nominative singular, the locative singular, and the instrumental singular. The nominative singular form of the noun is thoroughly independent of prepositions; there are none by which it is prefaced. In contrast, the locative singular depends solely and fully on a preposition for the specification of both meaning and case. There are six meanings associated with the locative singular and its orthographic form is not unique, since other grammatical cases of the noun are spelled the same way (for example, the dative singular). For each of the six locative singular meanings there is a preposition and that preposition necessarily and sufficiently specifies the meaning. The sentence context is superfluous. With regard to the instrumental singular case, it is in one sense simpler than the locative singular case, viz., there are no other cases with which it is orthographically identical. In another sense, however, the instrumental singular is more complex. It has sixteen possible meanings (Ivić, Note 1) where a meaning depends either on a preposition or on the sentence context. A preposition, therefore, is only occasionally necessary and sufficient to specify the meaning of a noun in the instrumental singular and is never needed to identify the case. To draw the contrast sharply: For a word in the locative singular an appropriate preposition indicates (1) that the word _is_ in the locative singular case and not in some other case (one that is spelled identically); and (2) which one of six potential meanings is to be ascribed to the locative singular. For a word in the instrumental singular, an appropriate preposition does not perform the role described in (1) but only a role similar to but weaker than that identified in (2).

One would intuit from the foregoing discussion that in everyday sentence comprehension an appropriate preposition would markedly facilitate, and an inappropriate preposition would likely hinder, the grammatical and semantic evaluation of a noun in the locative singular form. And in comparison, the positive contribution of an appropriate preposition to the evaluation of a noun in the instrumental singular form would be generally less marked, and the negative contribution of an inappropriate preposition would be negligible. Carrying this intuition over into the lexical decision task we would expect: (1) lexical decision to locative singular forms to be facilitated and inhibited by appropriate and inappropriate prepositional primes, respectively; (2) lexical decision to instrumental singular forms to be facilitated less and inhibited not at all by appropriate and inappropriate prepositional primes, respectively; and (3) lexical decision to nominative singular forms to be unaffected by prepositional primes of either kind.

METHOD

Subjects

Ninety-nine students from the Department of Psychology, University of Belgrade, received academic credit for participation in the experiment. A subject was assigned to one of nine subgroups, according to the subject's appearance at the laboratory, for a total of eleven subjects per subgroup.

Materials

Two types of slides were constructed. In one type, a string of Letraset lowercase Roman letters (Helvetia Light, 12 points) was arranged horizontally in the upper half of a 35-mm slide and in the other type, letters of the same kind were arranged horizontally in the lower half of a 35-mm slide. Letter strings in the first type of slide were always prepositions (or pseudoword analogues) and letter strings in the second type of slide were always inflected nouns (or pseudoword analogues). Altogether there were 120 "preposition" slides and 120 "inflected noun" slides with each set evenly divided into words and pseudowords. The 60 inflected noun slides that were words consisted of three sets of twenty representing the nouns, respectively, in nominative singular, locative singular, and instrumental singular. The twenty nouns were selected from the middle frequency range of a corpus of one million Serbo-Croatian words (Kostić, Note 2). A different set of twenty nouns of the same frequency was used to generate the pseudowords. This was done by simply changing the first letter of the nouns in the nominative singular and locative singular and by changing either the first letter or the final one or two letters for the nouns in instrumental singular.

Across genders the nominative singular form either ends in a vowel or a consonant, the locative singular always ends in a vowel, and the instrumental singular always ends in a consonant. Importantly, apart from the instrumental singular form and the occasional nominative singular form, the grammatical cases of Serbo-Croatian nouns end in a vowel. We wished to arrange matters so that both beginnings and endings of letter strings contributed to negative decisions. We also wished to do as little damage as possible to the root morphemes and to make the pseudoword versions of the nominative singular, locative singular, and instrumental singular cases of a given word form a coherent set. We would not substitute the vowel ending of a locative singular by another vowel ending because that would only generate the same word in another grammatical case. We could substitute another consonant for the terminal consonant of a nominative singular, but that would render the overall set of derived pseudowords less coherent than we desired because the nominative singular of nouns in the masculine is the root morpheme. We chose, therefore, to modify the endings of some of the nouns in instrumental singular. All things considered that seemed to us the most prudent manipulation.

The preposition slides and the inflected noun slides were grouped into pairs such that (1) the inflected noun slides contained a word in one half of the pairs and a pseudoword in the other half, and (2) the preposition slides contained a preposition specific to locative singular (one of na, po, pri), or a preposition specific to instrumental singular (one of sa, nad, pred), or a monosyllabic pseudoword (twelve pseudowords were used: uk, af, nu, fe, fo, pug, tir, dri, vak, knid, pler, tev). In total, there were 1,080 different pairs of slides, of which a given subject saw 120 pairs.

Design

As remarked, each word and pseudoword appeared in three grammatical cases. The major constraint on the design of the experiment was that a given subject never encountered a given word or pseudoword in any grammatical case more than once. This was achieved in the following manner.

Of the 120 word and pseudoword stimuli, 12 stimuli (six words and six pseudowords) were used for practice. The remaining 108 words and pseudowords were divided into three groups (A,B,C) with 36 items in each group. Each of these three groups was further divided into three subgroups (a,b,c) of 12 items each (six words and six pseudowords).

Ninety-nine subjects were divided into three groups (1,2,3) with 33 subjects in each group. Further division was undertaken where each group of subjects was divided into three subgroups (I,II,III) with 11 subjects each.

Note that there were six parameters in the design: three groups of words (A,B,C) with three subgroups each (a,b,c); three preposition types (locative-specific, instrumental-specific, and nonsense); three grammatical cases (nominative singular, locative singular, instrumental singular), and three groups of subjects (1,2,3), each divided into three subgroups (I,II,III). In short, each subject in each subgroup of eleven subjects saw each grammatical case-preposition type combination; but across the nine subgroups of eleven subjects, the nine grammatical case-preposition type combinations were defined on different subsets of twelve nouns (that is, six words and six pseudowords). Therefore, an individual subject, while seeing all grammatical case-preposition type combinations, never saw the same noun twice, but all subjects did see all 108 base stimuli. Put differently, each subject saw the same nouns as every other subject but not necessarily in the same grammatical case nor necessarily preceded by the same preposition type.

Procedure

Two slides were presented on each trial. The subject's task was to decide as rapidly as possible whether the letter string contained in a slide was a word or a pseudoword. Each slide was exposed in one channel of a three-channel tachistoscope (Scientific Prototype, Model GB) illuminated at 10.3 cd/m^2. Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches farther away. The closer button was depressed for a "No" response (the string of letters was not a word), and the farther button was depressed for a "Yes" response (the string of letters was a word).

Latency was measured from slide onset. The subject's response to the first slide terminated its duration and initiated the second slide unless the latency exceeded 1,300 msec, in which case the second slide was initiated automatically. The duration of the second slide, like that of the first, was terminated by the key press.

RESULTS AND DISCUSSION

Before considering the data of major interest, namely, the positive decision times for the noun targets, we give a brief summary of the decision times for the other letter-strings in the first and second lists of a pair. Average decision latencies for the pseudowords in nominative singular, locative singular, and instrumental singular were 711 msec, 706 msec, and 774 msec, respectively, when preceded by the instrumental prepositions; 713 msec,

726 msec, and 750 msec, respectively, when preceded by the locative prepositions, and 727 msec, 721 msec, and 784 msec, respectively, when preceded by the nonsense prepositions. The longer times for rejecting pseudowords in the instrumental singular are probably owing to their greater length (on average they were about one letter longer). The overall pattern of negative decision latencies for the three grammatical cases is similar to that reported by Lukatela, Mandić, Gligorijević, Kostić, Savić, and Turvey (1978). Importantly, regular and nonsense prepositions do not appear to have influenced decision times on pseudowords. With regard to the regular prepositions, the average latencies were 512 msec for the prepositions appropriate to the locative case and 514 msec for the prepositions appropriate to the instrumental case. The nonsense prepositions were rejected at an average latency of 682 msec.

Figure 1 presents mean positive decision times for each grammatical case and preposition. The figure is based on 52 words rather than the original 54; two words were aligned with the wrong prepositions and had to be discarded. Inspection of Figure 1 suggests that, as conjectured, preposition type did not affect decision times to nouns in the nominative singular, but did affect decision times to the same nouns in the locative singular and instrumental singular, particularly the former. This suggestion was substantiated by statistical analyses. In one analysis, a mean reaction time was computed for each subject by averaging over (approximately) six words (recall that two of the fifty-four words were discarded) in each combination of grammatical case and preposition type (locative specific, instrumental specific, and nonsense). An analysis of variance on these subjects' means revealed that preposition type was significant, $F(2,196)=18.9$, $MSe=62910$, $p < .001$, as was grammatical case, $F(2,196)=41.0$, $MSe=4904$, $p < .001$. Additionally, there was a significant interaction between grammatical case and preposition type: $F(4,392)=3.3$, $MSe=2027$, $p < .02$.

In another analysis, a mean reaction time was computed for each word by averaging over eleven subjects in each combination of grammatical case and preposition type. An analysis of variance on the means of these words revealed that preposition type and grammatical case were significant: $F(2,102)=10.66$, $MSe=29147$, $p < .001$ and $F(2,102)=28.19$, $MSe=3872$, $p < .001$, respectively. The interaction of preposition type and grammatical case, however, missed significance: $F(4,204)=1.51$, $MSe=3297$, $p > .05$.

Focusing now on the specific predictions, it was supposed that of the three forms the locative singular should be most affected by appropriate and inappropriate prepositions, the instrumental singular should be affected considerably less so and the nominative singular should not be affected at all. Inspection of Figure 1 confirms the predicted insensitivity of the nominative singular. T-tests computed over subjects and over words were used to compare the decision times to the locative singular form when that form was preceded by (1) a locative-specific preposition, (2) an instrumental-specific preposition, and (3) by nonsense. A comparison of (1) with (2) proved significant over subjects, $t(10)=6.27$, $p < .01$, and over words, $t(5)=4.20$, $p < .01$, as did a comparison of (1) with (3), $t(10)=4.27$, $p < .01$; $t(5)=2.87$, $p < .01$. A comparison of (2) with (3), however, revealed no significance either over subjects, $t(10)=1.99$, $p > .1$, or over words, $t(5)=1.34$, $p > .2$. Similar comparisons conducted for the instrumental singular form showed that the

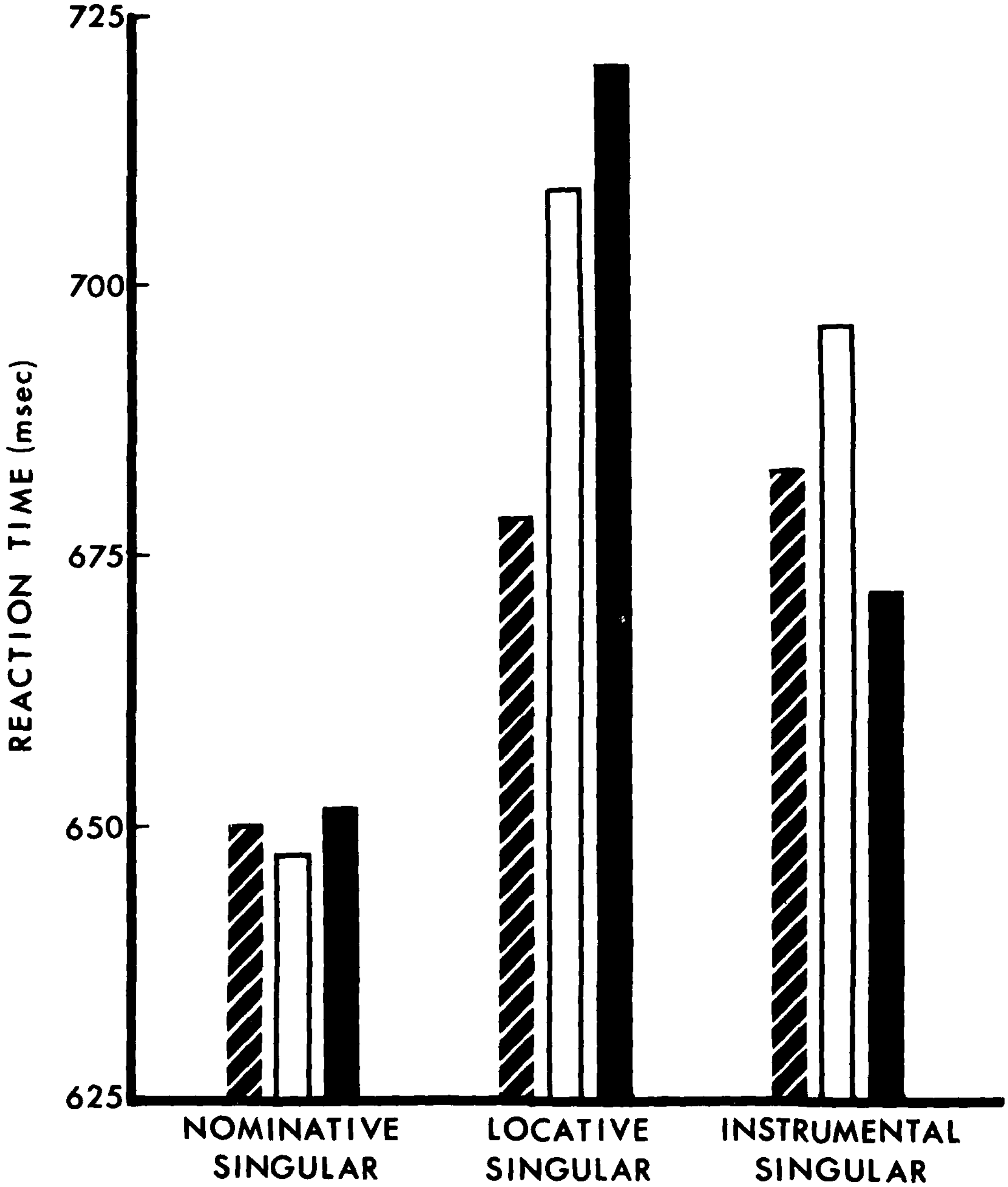


Figure 1. Positive lexical decision times for three grammatical cases preceded by locative-specific prepositions (striped), instrumental-specific prepositions (black) and nonsense (white) prepositions.

appropriate prepositions did not facilitate lexical decision in comparison to the inappropriate prepositions (over subjects, $t(10)=1.85$, $p > .1$; over words, $t(5)=1.24$, $p > .3$). There was evidence, however, that appropriate prepositions facilitated lexical decision in comparison to nonsense prepositions (over subjects, $t(10)=3.70$, $p < .01$; over words, $t(5)=2.49$, $p < .02$).

Lexical decision times were not always significantly slowed by inappropriate prepositions. Inspection of Figure 1 and the pattern of the t-tests reveal that the effect of inappropriate prepositions was not the same for the locative singular and the instrumental singular forms. Consistent with our suppositions, the data point to a detrimental effect of inappropriate prepositions on lexical decision only for the locative singular.

In sum, the results of the present experiment extend previous observations on the priming of the internal lexicon by demonstrating that such priming can occur for words that are not so much related semantically as they are related grammatically. Additionally, the outcome of the experiment lends support to the intuition that, in the reading of sentences, lexical facilitation occurs among sentential components.

REFERENCE NOTES

1. Ivić, M. *Case system in Serbo-Croatian language* (in Serbo-Croatian). Internal publication of the University of Novi Sad, 1964.
2. Kostić, Dj. *Frequency of occurrence of words in Serbo-Croatian*. Unpublished paper, Institute of Experimental Phonetics and Speech Pathology, Belgrade, Yugoslavia, 1965.

REFERENCES

Ammon, M. S., & Slobin, D. I. A cross-linguistic study of the processing of causative sentences. *Cognition*, 1979, *7*, 3-17.

Bradley, D. C. *Computational distinctions of vocabulary type*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1978.

Blank, M. A., & Foss, D. J. Semantic facilitation and lexical access during sentence processing. *Memory & Cognition*, 1978, *6*, 644-652.

Fischler, J. Semantic facilitation without association in a lexical decision task. *Memory & Cognition*, 1977, *5*, 335-339.

Forster, K. I. Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland, 1976.

Lukatela, G., Mandić, Z., Gligorijević, B., Kostić, A., Savić, M., & Turvey, M. T. Lexical decision for inflected nouns. *Language and Speech*, 1978, *21*, 166-173.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word recognition. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press, 1975.

Morton, J., & Long, J. Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 43-51.

Neely, J. H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 1977, 106, 226-254.

Patterson, K. E., & Marcel, A. J. Aphasia, dyslexia and the phonological coding of written words. Quarterly Journal of Experimental Psychology, 1977, 29, 307-318.

Scarborough, D. L., Cortese, C., & Scarborough, H. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 1-17.

Schuberth, R. F., & Eimas, P. Effects of context on the classification of words and pseudowords. Journal of Experimental Psychology: Human Perception and Performance, 1977, 33, 27-36.

Underwood, G. Contextual facilitation from attended and unattended messages. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 99-106.

AN EVALUATION OF THE "BASIC ORTHOGRAPHIC SYLLABIC STRUCTURE" IN A PHONOLOGICALLY SHALLOW ORTHOGRAPHY

Laurie B. Feldman, A. Kostić,+ G. Lukatela,++ and M. T. Turvey+++

Abstract. The notion of a "Basic Orthographic Syllabic Structure" (BOSS) (Taft, 1979a) was examined in the phonologically shallow orthography of Serbo-Croatian, which is a highly inflected language written in two alphabets--Roman and Cyrillic. Some characters are shared by both alphabets and retain the same pronunciation in each, some are unique to one alphabet, and some are ambiguous, i.e., receive different readings in the two alphabets. Thus, a letter string composed of common and ambiguous characters might be pronounced in one way if read in Roman and in a different way if read in Cyrillic. Lexical decisions were made on a set of words that met the following criteria: When written in Cyrillic, the nominative singular form of the word was phonologically ambiguous while the dative singular form of the word was unambiguous; when written in Roman, both grammatical forms of the word had only one possible pronunciation. The relation between the lexical decisions to the nominative singular and dative singular forms of the same word depended upon the alphabet in which the words were written. Decision times for the Cyrillic nominative singular forms were very slow relative to those for the Roman nominative singular; in contrast, the decision times for the Roman and the Cyrillic dative singular forms were virtually identical. The BOSS perspective anticipates the same relationship between grammatical forms in both alphabets, since inflected forms of the same word must share the same BOSS and their affixes must occur with the same frequency. In addition, the results showed that the number of ambiguous characters is a significant determinant of the decision latencies when no unique characters are present. The BOSS perspective was dismissed in favor of the view that the lexical representation of Serbo-Croatian words is phonological and not purely orthographic.

How does a reader determine that a string of letters is a word? The words a reader knows are said to be represented in a special memory conventionally termed the internal lexicon. Roughly, a representation is a structure whose elements (symbols, predicates, or whatever) putatively corres-

---

+Also University of Connecticut and University of Belgrade.
++University of Belgrade.
+++Also University of Connecticut.
Acknowledgment. This research was supported in part by NICHD Grant HD 08495 to the University of Belgrade, in part by NICHD Grant HD 01994 to Haskins Laboratories, and in part by NSF dissertation fellowship BNS 7924409 to Laurie Feldman.

[HASKINS LABORATORIES: Status Report on Speech Research SR-66 (1981)]

pond to the significant aspects of the thing represented; words are things that are heard, seen, spoken, and written; and to represent a word is to capture (aside from its semantic evaluation) the essential details of one or more (perhaps all) of its physical embodiments. Whatever might be the nature of such details, it is generally conceded that the vocabulary in terms of which a word is described in lexical memory will be nonidentical with the vocabulary in terms of which a word is described as a stimulus for a listener or viewer or as an activity of a speaker or writer. A notable example of this contrast of vocabularies is given in the contemporary analysis of speech perception: The descriptors of the acoustic embodiment of a word are dynamic, continuous, and context-dependent, whereas the descriptors of the word in memory are static, discrete, and context-independent. At all events, given distinct vocabularies, the answer to the above question, roughly speaking, is that the reader must internally translate the letter string into a vocabulary identical to that in which lexical entries are described so that the matching of stimulus and memory can be effected; if a match is made, the stimulus is a word. This brings us to the main question of the present paper: What is this proprietary vocabulary?

In response to this question students of reading have generally entertained two options: (1) that the proprietary vocabulary is one whose predicates are referential of the visual form of words; (2) that the proprietary vocabulary is one whose predicates are referential of the speech form of words. The first option could be pursued indifferent to any linguistic concerns. That is, one could imagine that the vocabulary consists of predicates that refer strictly to visual things--such as individual letter shapes, transgraphemic features, or a word's Fourier spectrum. An alternative tack is one in which the visually referential predicates are linguistically constrained. For example, if the predicates are referential of letter clusters, the letter clusters might conform to the morphology of the language. The predicate types, therefore, would refer to free stems, prefixes, inflections, and so on. Unlike the visual option, the speech option cannot relate arbitrarily to linguistic considerations. The predicates it mandates are referential of the significant phonological dimensions of speaking and of hearing speech--phonemes, featural decompositions of phonemes, syllables, etc.

In short, the two options emphasize two different physical embodiments of words: things produced by printing and writing and known by eye, and things produced by speaking and known by ear. Now it is of course a fact that all orthographies transcribe language and that all alphabetic orthographies are phonographic--they specify, more or less directly, how a word sounds. Nonetheless, arguments have been given for supposing that the proprietary vocabulary for describing the lexical entries of the fluent reader is not speech-related--at least not principally speech-related. The empirically based arguments have been ably reviewed (e.g., Coltheart, 1979). These arguments, by and large, are extensions and elaborations of an often-voiced claim that the linkage between the English orthography and the phonemic (and phonetic) structure it conveys is overly abstract (in the sense of involving many successive transformations) for the purposes of fluent reading. Given the expected difficulty (and, thus, the slowness) of recovering the (abstract) speech form of a word from its orthographic form, it has seemed better to assume that the lexicon's entries relate more closely to the written form than to the spoken form of the language.

Recently, Taft (1979a, 1979b) has characterized the English lexicon in terms of a predicate that is referential of both orthographic and morphologic factors. This predicate is termed the "Basic Orthographic Syllabic Structure" or "BOSS." Given a (nonprefixed) word, the BOSS is that part of the first morpheme that includes after its first vowel all consonants that do not violate rules of orthographic co-occurrence. BOSSes are said to be stored in a peripheral orthographic file that is distinguished from the main file in which all the information about a word is to be found.

The BOSS perspective exemplifies the class of visual options. It answers the question with which we began--of how a reader determines that a string of letters is a word--as follows: A presented word is first analyzed into affixes and stem, presumably by a procedure that refers to a lexical listing in which there are predicates referential of these morphemes. The accessing proper now begins in which a search is made of the orthographic file of successive letter groupings that begin with the first letter of the word (subsequent to any prefixes). Consider CANDLE as an example. The BOSS of CANDLE is CAND. The initial search of the orthographic file is for CA. This search would fail (that is, be exhaustive) and a second search would be initiated with CAN. This search would be successful but the specified address in the main file would prove to be inappropriate, precipitating yet a further search of the orthographic file--this time with CAND. Accessing CAND in the orthographic file would lead to the requisite entry in the main file where complete information on CANDLE is stored. In sum, whereas the representation of a word in memory is according to the BOSS principle, the means by which a word is retrieved is not; on the contrary, retrieval proceeds as a reiterative left-to-right search starting with the first letter of the root morpheme.

There is another aspect of lexical access to be remarked upon. BOSSes are arranged in the orthographic file according to their frequency of occurrence in the written language. Consequently, for two BOSSes of identical length (neither of which includes a word within itself, see Taft, 1979a), the BOSS that occurs more frequently will be found more rapidly. As noted, when a BOSS is found in the orthographic file, it gives an address in the main file. The word's stem and legal affixes in the main file are represented in a fashion that reflects the individual frequencies with which the stem and each of its given legal affixes co-occur. For two words with the same stem (and, therefore, the same BOSS) but with different affixes, the affixed form that is the more frequent will be detected in less time. According to the BOSS predicate view, the time taken to decide that a word is a word depends on both the frequency of the word's BOSS and the frequency of the word.

While Taft's principle for deriving lexical structure may be appropriate for the English orthography, it is unclear whether the principle is applicable to an orthography that is less distant from the (classical) phonemic (and phonetic) structures that it conveys, such as the orthography of Serbo-Croatian.

In contrast to English, which is morphophonemic in its referent (Chomsky, 1970), the writing system of Serbo-Croatian preserves a very close relation to (classical) phonemics and only reflects a common morphology when phonology is preserved. In Serbo-Croatian, all similar orthographic patterns will sound alike. Even fully systematic phonological alternation in surface forms is

represented in the orthography so that visual or orthographic similarity of morphologically related forms may be obscured, for example, nominative singular RUK+A, dative singular RUC+I; nominative singular SNAH+A, dative singular SNAS+I. (Note: Inflection is the major grammatical device of Serbo-Croatian and the preceding are Roman transcriptions [see below] of the English words arm and daughter-in-law, respectively.) In addition, as a result of the tendency toward open syllables in Serbo-Croatian, the possible patterning of consonants and vowels is much more restricted in Serbo-Croatian than in English. Not only do the orthotactic (Taft, 1979a) rules fully mimic the phonotactic rules, but the possibility for ambiguous syllable boundaries due to sequences of consonants is greatly reduced.

In sum, the Serbo-Croatian orthography relative to the English orthography permits less variability in its orthographic patterning, is more closely related to the spoken language, and is less concerned with preserving morphological invariance. Collectively, the inference is that BOSSes are less likely to be elemental predicates in the proprietary (internal) vocabulary of Serbo-Croatian and this will be evaluated in the present experiment.

Serbo-Croatian is written in two alphabets, Roman and Cyrillic, both of which were constructed in the last century according to the simple rule: "Write as you speak and speak as it is written." Both the Roman and Cyrillic orthographies transcribe the sounds of the Serbo-Croatian language in a regular and straightforward fashion, and there are no (nontrivial) derivation rules to speak of.

The Roman and Cyrillic alphabets map onto the same set of phonemes but comprise two sets of letters that are, with certain exceptions, mutually exclusive (see Figure 1 and Table 1). Most of the Roman and Cyrillic letters are unique to their respective alphabets. There are, however, a number of letters that the two alphabets have in common. The phonemic interpretation of some of these shared letters is the same whether they are read as Cyrillic or as Roman letters; these are referred to as <u>common</u> letters. Other members of the shared letters have distinct phonemic values in Roman reading and in Cyrillic; these are referred to as <u>ambiguous</u> letters. Within each category, the individual letters of the two alphabets have phonemic values that are virtually invariant over letter contexts. Moreover, all the individual letters in a string of letters, be it a word or nonsense, are always pronounced--there are no letters made silent by context. Finally, but not least in importance, we should note that a large portion of the population uses both alphabets competently. This is due, in part, to an educational requirement that both alphabets be taught within the first two grades. Roman is taught first in the western part of Yugoslavia and Cyrillic in the eastern part of Yugoslavia.

Given the nature of and the relation between the two Serbo-Croatian alphabets, it is possible to construct a variety of types of letter strings. A letter string of uniquely Roman letters or of uniquely Cyrillic letters would be read in only one way and could be either a word or nonsense. A letter string composed of the common and ambiguous letters could be pronounced one way if read as Roman and pronounced in a distinctively different way if read as Cyrillic; moreover, it could be a word in one alphabet and nonsense in the other, or it could represent two different words, one in one alphabet and one in the other, or it could be nonsense in both alphabets.

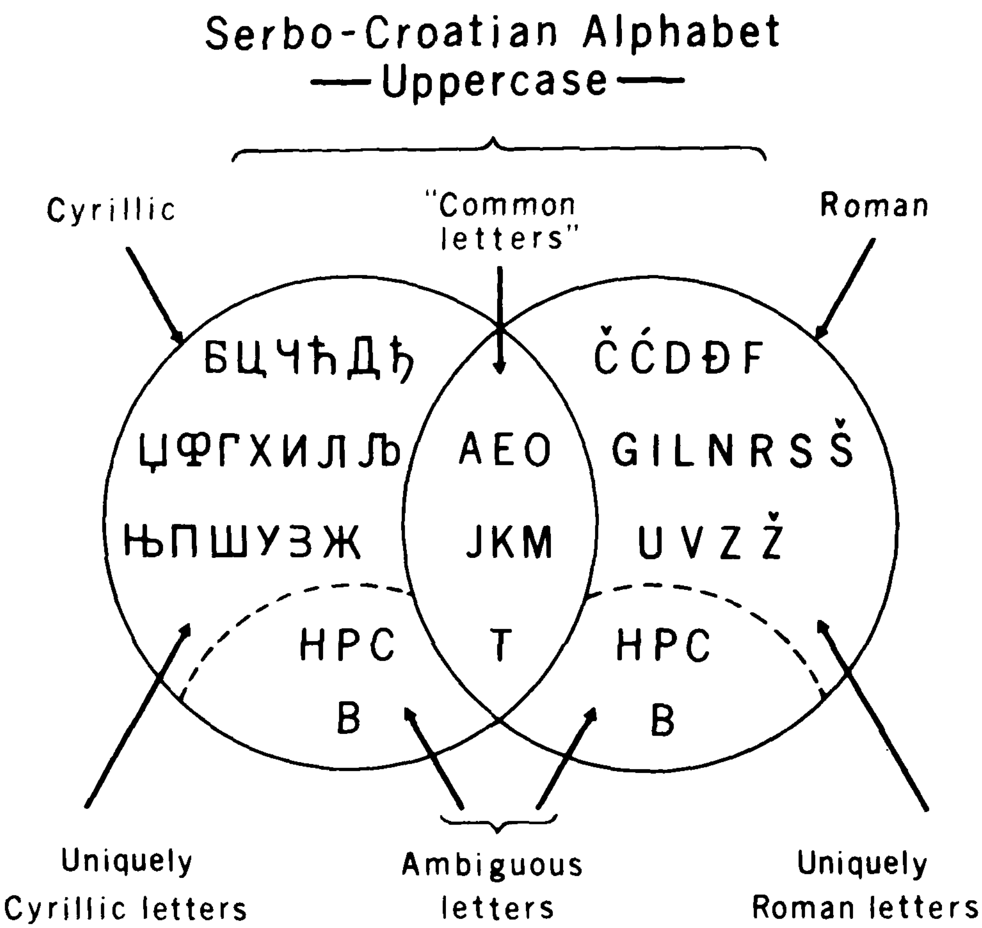


Figure 1. The uppercase characters of the Roman and Cyrillic alphabets of Serbo-Croatian.

TABLE 1

SERBO-CROATIAN				
ROMAN		CYRILLIC		
PRINTED		PRINTED		LETTER NAME IN I.P.A.
UPPER CASE	LOWER CASE	UPPER CASE	LOWER CASE	
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	nə
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

Consider letter strings of the following type: VENA and BEHA, TONA and TOHA. The first letter string of each pair is the nominative singular form of a noun (English vein for the first pair and tone for the second pair) written in its Roman form and the second letter string of each pair is the same grammatical case of the same noun as it is written in its Cyrillic form. The Roman form of both pairs is written in a mixture of common letters and uniquely Roman letters, whereas the Cyrillic form of both pairs is a mixture of common letters and ambiguous letters (two in the Cyrillic member of the first pair and one in the Cyrillic member of the second pair). Importantly, the Cyrillic form of each pair contains no unique (Cyrillic) letters--that is, nothing that marks it as a letter string to be read specifically in one alphabet or the other; additionally, the Cyrillic form of each pair is nonsense if given a Roman reading. Let us now extend the above short list of letter strings to include their respective dative singular cases: VENA, VENI: BEHA, ВЕНИ: TONA, TONI: TOHA, ТОНИ. What is important to note here is that in the dative case, the Cyrillic form now includes a uniquely Cyrillic character that would specify the particular alphabet in which the letter string is to be read. Table 2 summarizes the foregoing contrasts.

Table 2

Examples of Serbo-Croatian Words in Two Grammatical Cases: Written in Two Alphabets

Meaning	Alphabetic Transcription	Nominative Singular	Dative Singular
Tone	Cyrillic	TOHA	ТОНИ
	Roman	TONA	TONI
Vein	Cyrillic	BEHA	ВЕНИ
	Roman	VENA	VENI

The present experiment looks at the following special version of the question asked at the outset: How does a bi-alphabetical reader of Serbo-Croatian determine that a letter string of the kind depicted in Table 2 is a word? We identify below the five hypothetical answers to this question, together with the particular predictions that follow from them. Four of these

hypotheses assume that the proprietary vocabulary for describing access to the lexical representations of Serbo-Croatian words is orthographic. More precisely, these hypotheses derive from the BOSS perspective, with the first two adhering strictly to Taft's (1979a, 1979b) original formulation and with the third and fourth being modifications of that formulation to accommodate the two-alphabet nature of the Serbo-Croatian orthography. The fifth hypothesis contrasts with the previous four in that it assumes that the lexical representations of Serbo-Croatian words are written in a speech-related vocabulary. This fifth hypothesis follows, in part, from a consideration of the design of the Serbo-Croatian orthography and, in part, from the data of various lexical decision experiments conducted with the orthography (Feldman, 1980; Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978; Lukatela, Popadić, Ognjenović, & Turvey, 1980).

1. <u>The Roman-bias Hypothesis</u>

In a list of words respecting the contrasts of Roman and Cyrillic forms identified above, <u>only</u> the Roman forms are always unambiguous; and in the experiment to be reported that examines such lists, three quarters of the presented stimuli are in the Roman forms. The first hypothesis underscores this Roman bias in the materials by assuming that it similarly characterizes the readers themselves. That is, it makes the assumption that the readers, in their past, have primarily (but far from exclusively) encountered the Serbo-Croatian language transcribed in Roman. We would expect, therefore, that words formed from Roman BOSSes will generally incur shorter search times than the equivalent Cyrillic BOSSes--VENA, VENI, TONA, TONI should be associated with shorter lexical decision times than BEHA, BEHИ, TOHA, TOHИ, respectively. Moreover, because the declension affixes for the Roman and for the Cyrillic forms of the same word must relate among themselves in the same way (with regard to their relative degree of attachment to the stem), we would expect that the decision latencies for VENA and BEHA, and TONA and TOHA, will be less than those for VENI and BEHИ, TONI and TOHИ. This latter prediction is based on the fact that the nominative singular for any given Serbo-Croatian noun occurs much more frequently than the dative singular (Dj. Kostić, Note 1; Lukatela, Gligorijević, A. Kostić, & Turvey, 1980).[1] Finally, by the present hypothesis, latencies should not depend in any way on the number of ambiguous characters.

2. <u>The Non Alphabet-bias Hypothesis</u>

The assumption here is that the reader has experienced the Serbo-Croatian language equally in the two alphabets. Thus the overall frequencies with which the orthographic forms BEHA and VENA, BEHИ and VENI, TOHA and TONA, TOHИ and TONI have been experienced will be <u>at least</u> equal. But it may well be the case that the overall frequency of the Cyrillic stems (and BOSSes), for example, BEH, will be greater than the overall frequency of the Roman stems (and BOSSes), for example, VEN, because BEH is the orthographic form not only of /bexa/ in Cyrillic but also of /vena/ in Roman, whereas VEN is the orthographic form only of /vena/. Thus the search time for the BOSSes of the Cyrillic forms and of the Roman forms of the same word will either be equal or different in favor of the Cyrillic forms. As with the previous hypothesis, however, the latencies for nominative singular cases should be shorter than for their respective dative singulars and the number of ambiguous characters should be irrelevant.

3. <u>Two Distinct Orthographic Files: The Parallel Search Hypothesis</u>

Let us assume that there are <u>two</u> orthographic files, one for Roman BOSSes and one for Cyrillic BOSSes. An individual bi-alphabetical reader might have more experience with the BOSSes of one alphabet than with those of the other, but this should not alter the relative orderings of BOSSes within the two files; that is, the BOSS of VENA (viz., VEN) and the BOSS of BEHA (viz., BEH) should be located in exactly the same places in the Roman and Cyrillic orthographic files, respectively (even though in the Roman file, VEN and BEH may occupy very different locations). Similarly, the relative frequencies with which inflected endings are affixed to stems in the main file should not differ; that is, there should be no difference between the relative attachments of A and I to VEN and the relative attachments of A and И to BEH. Let us now consider that the two files are accesssed in parallel. For any given letter string, the inflected ending is stripped off and the left-to-right reiterative retrieval procedure is conducted simultaneously in both orthographic files. Thus, VENA would be parsed into VEN + A and the retrieval would proceed first with VE (unsuccessfully in both files) then with VEN (unsuccessfully in the Cyrillic files, successfully in the Roman file). Similarly BEHA would be parsed into BEH + A and the retrieval would proceed first with BE (unsuccessfully in both files) then with BEH (possibly successfully in both files but always faster in the Cyrillic file). Given that BOSSes of the same Serbo-Croatian word are located at virtually identical sites in the two files, the times to find VEN and TON in the Roman file should be roughly equal to the times to find BEH and TOH, respectively, in the Cyrillic file. And, likewise, the time to confirm the legality of the BOSS and affix combination should be roughly equal for the Roman and Cyrillic transcriptions. Thus, by the present hypothesis, lexical decision times to the Cyrillic and Roman transcriptions of the same word in the nominative singular (BEHA, VENA and TOHA, TONA) should not differ; nor should lexical decision times to the Cyrillic and Roman transcriptions of the same word in the dative singular (BEHИ, VENI and TOHИ, TONI); but as with the previous two hypotheses, decision times should be shorter to the nominative singular case of a word than to the dative singular case and the number of ambiguous characters should not be a determinant of decision times in either grammatical case.

4. <u>Two Distinct Orthographic Files: The Successive Search Hypothesis</u>

There are two versions of this hypothesis because there are two stages prior to retrieval proper that must be proposed--parsing and alphabet determination--and the predictions differ depending on how the two stages are ordered. Let the parsing occur first. Then, having removed the grammatical affix from the stem, a search is made of the stem to determine whether it includes a unique character. If the search is positive, then the first unique character found is evaluated for its alphabet status: if it is Roman, the search for the appropriate BOSS unit is directed to the Roman orthographic file; if it is Cyrillic, the search for the appropriate BOSS unit is directed to the Cyrillic orthographic file. However, if no unique character is found in the stem, then the choice whether to direct the search for the appropriate BOSS unit to the Roman file or to the Cyrillic file is random. Thus, whereas a stem such as VEN specifies its file (viz., Roman), a stem such as BEH does not. Therefore, on average, on half of the times that they occur, letter

strings such as BEHA, BEHИ, TOHA, TOHИ, will involve a successive search of both files, so that overall the left-to-right BOSS search and associated decision latency will be slower than the left-to-right BOSS search and decision latency associated with letter strings such as VENA, VENI, TONA, TONI. There are, therefore, two predictions of the parsing-first successive search hypothesis: one prediction is the same as that for parallel search, namely, that the latency difference between grammatical cases of the same word should not differ as a function of the alphabet in which the word is written; the other prediction is that the decision latency for a word transcribed in Roman should be less than the decision latency for the same word transcribed in Cyrillic.

Assume now that alphabet determination precedes parsing. This means that BEHA and TOHA will be treated differently from BEHИ and TOHИ. The first stage will determine that the BOSSes of BEHИ and TOHИ, isolated in the next and parsing stage, are to be searched for in the Cyrillic orthographic file; as before, however, where the BOSSes of BEHA and TOHA are to be found remains ambiguous. This variant of successive search makes a very different prediction from either the parallel search hypothesis or the parsing-first, successive search hypothesis: it predicts that the lexical decisions on BEHA and TOHA should be <u>slower</u>, respectively, than the lexical decisions on BEHИ and TOHИ; and that the lexical decisions on BEHИ and TOHИ should <u>not</u> differ from the lexical decisions on their Roman equivalents, VENI and TONI. It also predicts, consonant with each of the preceding hypotheses, that VENA will be faster than VENI, and TONA faster than TONI; and that the number of ambiguous characters is irrelevant to lexical decision.

5. <u>The Speech-related Hypothesis</u>

In this last hypothesis, the previously assumed orthographic basis of the lexicon is dismissed in favor of the assumption that the lexical representation of Serbo-Croatian words is phonological. Therefore, any given letter string must be encoded "phonemically" to effect a lexical search and a possible match, and this is achieved presumably by the transparent correspondences that define the orthography's relation to the phonemes of the language. The ambiguous characters are an exception of sorts to this transparency. In the absence of a unique character in a string of letters, any ambiguous character is necessarily equivocal with respect to the phonemic reading it will be given. Let us assume here, as we did with the previous hypothesis, that there is a preceding stage of alphabet determination. The detection of a unique character and of its alphabetic allegiance identifies the requisite set of grapheme-to-phoneme correspondences to be applied to the ambiguous characters. (We are not yet convinced that this is the best way of expressing the means by which ambiguous characters are disambiguated, but it will suffice for our present purposes.) For a letter string such as BEHИ, therefore, the presence of И specifies that B is to be read as /v/ and H is to be read as /n/; thus BEHИ (and, of course, VENA, VENI, TOHИ, TONI, TONA) would receive a unique phonemic transcription and, generally speaking, entail a single search of the lexicon. (As is conventional, search time is conceived as an inverse function of a word's frequency.)

In contrast, BEHA, which has no unique characters, can be transcribed phonemically in more than one way and could, therefore, involve more than one

search of the lexicon. Importantly, it is assumed that the assignment of a phoneme to an individual character in a letter string is a process that occurs independently of the assignment of phonemes to its neighbors; more fundamentally, it is a process that operates without knowledge as to the alphabet "rationalizing" any individual phonemic interpretation. Thus BEHA can be transcribed phonemically as /bena/, /vexa/, /bexa/ and /vena/, and if lexical search is with respect to one such phonemic transcription at a time, BEHA could entail, in principle, four successive searches of the lexicon until a match is found (with /vena/). (But see Lukatela, Popadić, Ognjenović, & Turvey [1980] for a parallel-search interpretation consonant with Morton's [1969] logogen theory.) Words with <u>two</u> ambiguous characters and no unique characters would contrast, by the foregoing argument, with words with <u>one</u> ambiguous character and no unique characters. TOHA can be ascribed only <u>two</u> phonemic readings--/toxa/ and /tona/--and, therefore, should entail at most two successive searches of the lexicon. In sum, by the present hypothesis: (1) the lexical decision times for BEHA and TOHA should be respectively <u>longer</u> than the lexical decision times for BEHИ and TOHИ; (2) the lexical decision times for VENA and TONA should be respectively <u>shorter</u> than the lexical decision times for VENI and TONI (by the standard argument based on the different frequencies of the two grammatical cases); (3) the lexical decision times for TOHИ and TONI should not differ nor should the lexical decision times for BEHИ and VENI; and (4) the lexical decision times for BEHA relative to VENA should be <u>longer</u> than the lexical decision time for TOHA relative to TONA.

<u>METHOD</u>

<u>Subjects</u>

Sixty-eight first year students of psychology at the University of Belgrade participated in this experiment in partial fulfillment of course requirements. Eight subjects' data were eliminated from the statistical analysis because their error rate on the critical test stimuli exceeded 40%. As there were only seven such stimuli to which the criterion for eliminating subjects was applied, 40% corresponds to missing only three items. The <u>overall</u> error rate proved to be extremely low--less than 1%.

<u>Stimuli</u>

All stimuli in the experiment consisted of letter strings that contained four characters patterned as CVCV. Each of the word stimuli was a noun and each of the pseudoword stimuli was derived by changing one or two letters in a (different) CVCV word. Consonant with the examples of Table 1, seven words were chosen (Set A), which in the nominative singular case, written in the Cyrillic form, contained only those letter strings shared by both alphabets. As a result, these letter strings that are words in Cyrillic can also be read as pseudowords in Roman, e.g., TOHA can be /tona/, a word, or /toxa/, a pseudoword. Four of these words had two ambiguous letters and two common letters and three of these words contained one ambiguous letter and three common letters. In their Roman transcription, all of these words contained at least one unique letter. In contrast to the nominative singular declension ending, the dative singular ending will always uniquely specify the appropri-

ate alphabet. The dative singular form for words presented in this experiment requires either И or У in Cyrillic or their equivalent, I or U, in Roman. All four of these characters are unique to one alphabet. For these words (Set A), alphabetic ambiguity occurs in the Cyrillic nomintaive singular. It is resolved in the dative singular form and it never occurs in the Roman versions of the same word.

Another group of seven words with CVCV pattern (Set B) was also presented in Roman and in Cyrillic and in the nominative and the dative singular declensions. In contrast to the Set A words, these words contained unique letters in both declined forms of both alphabetic transcriptions; in short, no letter string in the Set B stimuli was ever ambiguous.

It should be underscored that the small size--seven--of the critical word Set A (and therefore of its control, Set B) is a necessary consequence of the criteria that had to be met in order to produce the kinds of contrasts between Cyrillic and Roman forms of the same words that the experimental hypotheses required.

In the experiment, four groups of subjects saw some form of the same 28 words and 28 pseudowords on which they performed a lexical decision judgment. The two sets of experimental words were each presented in complementary combinations of nominative/dative and Cyrillic/Roman to the four groups of subjects. If Set A words were presented in Roman dative singular form to one group of subjects, then Set B words were presented to that same group in Cyrillic nominative singular form. In addition, all four groups saw the same seven words that could be read in the same way in either Cyrillic or Roman (common words) and the same seven words that could be read only in Roman. The pseudoword set, constant across the four subject groups, consisted of seven Roman (pseudo) dative singular, seven Cyrillic (pseudo) nominative singular and fourteen Roman (pseudo) nominative singular forms. This variability was introduced in order to make the pseudowords analogous to the word forms.

In summary, each of four groups of fifteen subjects saw seven words in dative singular word form, seven Cyrillic words, seven common words, and seven Roman words, as well as seven Cyrillic and fourteen Roman pseudowords in nominative singular form and seven Roman pseudowords in dative singular form. Set A and B both appeared (between subject groups) in all four combinations of Roman/Cyrillic and nominative/dative, but these two sets differed in one important respect: The nominative Cyrillic form of Set A words contained only common and ambiguous letters. As a result, these strings, which are words in Cyrillic, can also be read as Roman pseudowords, e.g., MEPA can be /mera/ or /mepa/. Note that this alphabetic ambiguity is resolved in the dative singular form of these words, e.g., МЕРИ, and never occurs in the Roman version of the same word. By contrast, all forms of the Set B words are always unambiguous in their reading. That is, the words include unique characters in both nominative singular and in dative singular, for both the Roman and Cyrillic transcriptions (e.g., ЖАБА, ŽABA, ЖАБИ, ŽABI).

<u>Procedure</u>

In the instructions to the subject that preceded the experimental session, the variety of stimulus forms (nominative/dative singular, Cyrillic/Roman) was noted.

Each stimulus was presented for 500 msec in one field of a scientific Prototype Model GB tachistoscope and reaction time was measured from a counter that began with the stimulus onset. The blank field preceded the presentation of each stimulus and reappeared immediately after each response. The inter-stimulus interval was about 3 seconds and a short practice session preceded the experiment. All stimuli were typed on Prima U film and Cyrillic and Roman typeface were closely matched for size and form. (Common letters were identical in the two typefaces.)

Subjects performed a lexical decision task and tapped one of two telegraph keys. They depressed the closer key (thumbs) if the letter string was a pseudoword and the further key (forefingers) if the letter string was a word. Subjects were informed by the experimenter if they made an error on one of the test stimuli. A practice sequence of eight items preceded the experimental session.

RESULTS

An analysis of variance performed on all stimuli revealed no significant difference between the four groups of subjects, $F(3,56) = 0.13$, but significant main effects of lexicality (word-pseudoword), $F(1,56) = 123.9$, $MS_e = 11981$, $p < .001$, and word set, $F(3,168) = 82.9$, $MS_e = 3544$, $p < .001$. The word-set-by-experimental-group interaction was significant, $F(9,168) = 12.43$, $MS_e = 3544$, $p < .001$, as were the word-set-by-lexicality and the word-set-by-lexicality-by-group interactions, $F(3,168) = 99.6$, $MS_e = 2533$, $p < .001$, and $F(9,168) = 18.1$, $MS_e = 2533$, $p < .001$, respectively. Mean latencies for types of words were 795 (averaged over all forms for Set A ambiguous words), 708 (for all forms of Set B unambiguous words), 616 (for common words), and 630 (for Roman nominative controls). For the pseudowords, mean latencies were 769 Roman pseudo datives), 870 (Cyrillic pseudo nominative), and 778 (for Roman pseudo nominative controls).

Two subsequent analyses of variance were performed including (1) only the four forms of the words in critical Set A and (2) only the four forms of the words is critical Set B. Table 3 summarizes the data for Set B and Table 4 summarizes the data for Set A. In these two tables, alphabet (Roman/Cyrillic) and case (nominative/dative singular) combine to define the four groups of subjects who saw different forms of the same seven words. For Set B (words chosen so as to contain unique letters both in Roman and in Cyrillic), Roman alphabet is faster than Cyrillic, $F(1,56) = 4.58$, $MS_e = 12715$, $p < .05$, and nominative case is faster than dative, $F(1,56) = 11.0$, $MS_e = 12715$, $p < .005$. (This is consistent with Lukatela et al., 1978; Lukatela, Gligorijević, Kostić, & Turvey, 1980.) There is no alphabet-by-case interaction, $F(1,56) = 0.44$.

The Set A (ambiguous Cyrillic form) words present a very different pattern, however. Here again, the main effects of case and alphabet are significant, $F(1,56) = 4.60$, $MS_e = 12565$, $p < .05$ and $F(1,56) = 22.95$, $MS_e = 12505$, $p < .001$, respectively. In addition, the case by alphabet interaction is significant, $F(1,56) = 29.25$, $MS_e = 12565$, $p < .001$. An examination of means by protected t-tests revealed no difference for Cyrillic and Roman versions of the dative singular case, and a very significant difference

Table 3

Mean Lexical Decision Response Latencies for Unambiguous Words (Set B)

	Nominative	Dative
Cyrillic	701	778
Roman	619	735

Table 4

Mean Lexical Decision Response Latencies for Ambiguous Words (Set A)

	Nominative	Dative
Cyrillic	921	805
Roman	617	833

between the Cyrillic and Roman nominative singular forms, $t(14) = .44$, $p < 1$, and $t(14) = 7.2$, $p < .01$, respectively. Relative to the Roman nominative singular and to both Roman and Cyrillic dative singular forms, the Cyrillic nominative singular is slow.

Finally, a t-test was conducted on the difference between a subject's mean latency for words with one and words with two ambiguous letters as compared with the same difference for the unambiguous forms of the same words. This test (with one subject deleted due to excessively long latencies relative to his mean reaction time) revealed that the degree of impairment due to phonological ambiguity, that is, the difference between the Roman nominative and the Cyrillic nominative singular forms, depends on the number of ambiguous letters, $t(27) = 2.70$, $p < .05$ (See Table 5).

Table 5

Mean Latencies for Lexical Decision to Words with One and with Two Ambiguous Letters in Their Cyrillic Form as Compared with the Roman Form of the Same Word

Number of Ambiguous Characters	Alphabet Transcription	Nominative Singular		Difference Between Nominative Singulars	Dative Singular		Difference Between Nominative Singulars and Dative Singulars
1	Cyrillic	ТОНА	862	229	ТОНИ	815	47
(unambiguous control)	Roman	TONA	633		TONI	855	-222
2	Cyrillic	ВЕНА	979	379	ВЕНИ	794	185
(unambiguous control)	Roman	VENA	600		VENI	811	-211

DISCUSSION

In the introduction, five hypotheses were identified that mapped the word forms in Table 2 onto a pattern of lexical decision times. The first two hypotheses assumed that BOSSes of Serbo-Croatian words were stored indifferent to alphabet in a single orthographic file. The fundamental prediction of these two hypotheses was that a latency difference between the nominative

AD-A105 090 HASKINS LABS INC NEW HAVEN CT F/G 17/2
STATUS REPORT ON SPEECH RESEARCH, 1 APRIL-30 JUNE 1981.(U)
1981
UNCLASSIFIED SR-66(1981) PHS-N01-HD-1-2420
NL

4 OF 4
AD A
105090

END
DATE
FILMED
10-81
DTIC

singular and the dative singular cases of the same word should not differ as a function of the alphabet in which the word is written. Inspection of Table 3 and the allied analysis of variance verify this prediction for the words of Set B, which were composed solely from common and unique characters in either the Roman or Cyrillic transcription. The prediction, however, is not confirmed for the words of Set A (words that are exemplified in Table 4), which, when written in Cyrillic, are composed of common and ambiguous characters in the nominative singular case and of common, ambiguous and unique characters in the dative singular case; and which, when written in Roman, are written solely in common and unique characters for both cases. For words of this latter kind, latencies for the Cyrillic transcription and for the Roman transcription of the nominative singular case were, respectively, significantly longer and significantly shorter than the latencies for their dative singular equivalents. This interaction can be seen in Tables 4 and 5 and was verified by the analysis of variance and protected t-tests. We therefore reject the first two hypotheses, that is, the hypotheses that follow almost directly from the relation among entries formulated by Taft (1979a, 1979b).

The third and fourth hypotheses adhered to the conceptions of the BOSS unit and the orthographic file, but allowed that there might be two orthographic files--one for the Cyrillic transcription of words and one for the Roman transcription of words. On the assumption that these two files could be searched in parallel, it was predicted that the Cyrillic and Roman transcriptions of the same word in the same grammatical case would be associated with the same decision latency. This prediction was not confirmed, which, of itself, is not a very serious indictment of the hypothesis. The analysis of both Set A and Set B words revealed an alphabet difference: Roman words were generally responded to faster than Cyrillic words. A variety of reasons can be given for the Roman superiority that would not impugn the hypothesis. For example, perhaps the feature set of Cyrillic characters is less compact than its Roman equivalent and therefore encoded with greater difficulty; or, that the subjects of the experiment were more facile at searching the Roman file. Of larger significance is the failure of the prediction that the parallel search hypothesis shares with the first two hypotheses, namely, that the various grammatical cases of the same word should be organized in the same way when transcribed by the Roman and Cyrillic alphabets. Again, Set B words confirmed the prediction but the critical words, those of Set A, gave strong evidence of an alphabet-induced interaction. The parallel search (of two orthographic files) hypothesis is therefore rejected.

The fourth hypothesis, which assumed a successive search of the two orthographic files, took two forms. The parsing-first form can be rejected for the same reason that we have rejected the first three hypotheses--because, like them, it predicts a non-interaction for Set A words with alphabet. Additionally, but less importantly, it can be rejected because it predicts that for Set A words, all Roman transcriptions would be associated with shorter decision latencies than their Cyrillic equivalents. This was not so for the dative case. The parsing-second version of the successive search hypothesis is, however, much less easily dismissed. It successfully predicts the alphabet-dependent relation of grammatical cases that was observed for Set A words and it successfully predicts (but, again, of lesser importance) the absence of a difference between Roman and Cyrillic transcriptions of the dative singular case of Set A words. Of course, it also predicts the pattern

of latencies for Set B words. What the parsing-second version of the successive search hypothesis does not predict, in concert with the first three hypotheses, is that the number of ambiguous characters in the Cyrillic transcription of a Set A word should make a difference.

Let us now consider the fifth hypothesis, which departs from the other four in that it assumes a phonological vocabulary for describing lexical entries rather than an orthographic vocabulary. This hypothesis predicted the interaction observed in the Set A words, the absence of a difference between Roman and Cyrillic transcriptions of the dative singular case of Set A words, and that the number of ambiguous characters should significantly affect lexical decision on words that are written only in common and ambiguous characters. Finally, congruent with the four preceding hypotheses, it predicted the results for Set B words, viz., that the nominative singular of a word should be responded to faster than the dative singular of the same word when those words, in either Roman or Cyrillic form, are not solely composed of common and ambiguous characters.

Patently, only the speech-related hypothesis and the parsing-second, successive search hypothesis (the former emphasizing phonology and the latter emphasizing orthography) emerge as potential answers to the question of how a bi-alphabetical reader of Serbo-Croatian determines that a letter string is a word. The two hypotheses are distinguished in the data of the present experiment by one fact: That two ambiguous characters slow lexical decision more than one ambiguous character slows lexical decision when there are no unique characters to resolve the ambiguity. This fact is predicted by the speech-related hypothesis but not by the successive-search hypothesis. Admittedly, resolution of theoretical issues in science sometimes turns on "small" empirical findings. Is there license to assume that the present "small" finding, a difference established on seven words, is one to which we can grant such status? The reader is reminded that the seven words of the critical set, Set A, probably constitute a majority of the words that meet the criteria needed to evaluate the hypotheses. Moreover, the difference under consideration is within-words: it is a difference between two values, each of which is a measure of the degree to which a word transcribed in Cyrillic differs from itself transcribed in Roman. Therefore the comparison of the difference between BEHA and VENA and the difference between TOHA and TONA is not contaminated by variability in word frequency, orthographic regularity, pronounceability, etc. All the standard confounding factors are removed by taking the difference between a word and itself as the unit of comparison; and yet the latency difference under consideration is of the order of 150 msec (see Table 5). To these points we add that in another experiment that has looked more generally at the influence of number of ambiguous characters, significant effects have been found. Two and three syllable words written solely in common and ambiguous characters were compared with themselves; that is, with the same word written solely in common and unique characters. The lexical decision times for the two syllable words differed by 255 msec for one ambiguous character and by 325 msec for two ambiguous characters. Similarly, the lexical decision times for the three syllable words differed by 245 sec for two ambiguous characters and by 349 msec for three ambiguous characters (Feldman, 1980, 1981). In sum, it seems fair to conclude that the number of ambiguous characters in a word that has no unique characters is a significant determinant of the time required to evaluate the word's lexical status.

It would be a mistake, however, to focus on the significance of the *number of ambiguous characters to the detriment of* the observation that the relation among the nominative singular and dative singular cases of Set A words was alphabet-dependent. That observation is sufficient to disarm a BOSS/orthographic file interpretation of the Serbo-Croatian (internal) lexicon. Only a very special concession, viz., that there are two orthographic *files, each of which is sensitive to the alphabet determination of* any grammatical affix, makes the observation on the number of ambiguous characters critical. While it is possible to interpret the present data with respect to a successive search of two orthographic files, each of which is effectively organized in a different fashion, this successive search interpretation would *not hold for previous results. Pseudowords composed of entirely common* letters (that are alphabetically bivalent but phonologically unique) were no slower than pseudowords containing unique letters (Lukatela, Popadić, Ognjenović, & Turvey, 1980).

All things considered, the present *experiment is consistent with the* claim that word recognition in Serbo-Croatian is necessarily phonological and further, it extends that claim. In previous experiments, a between-words effect of phonologically bivalent letter strings was assessed relative to different letter strings (Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela et al., 1978) and a within-words *effect of bivalent phonology was* demonstrated relative to an unambiguous transcription of the same letter string (Feldman, 1980, 1981). In the present experiment, phonologically ambiguous BOSS units were evaluated relative to the unique alphabet transcription of the same BOSS. Results indicate that the effect of bivalence was obtained only when the BOSS unit as well as its *grammatical affix* were ambiguous.

How then does a reader determine that a string of letters is a word? For the Serbo-Croatian orthography we wish to conclude that he or she does so by encoding the written word into an internal speech-related *vocabulary; in* short, we conclude that the proprietary vocabulary for the internal lexicon in Serbo-Croatian is phonological.

REFERENCE NOTE

1. Kostić, Dj. The structure of usage value of grammatical forms in Serbo-Croatian. Report of the Institute of Experimental Phonetics and Speech Pathology, Belgrade, 1965.

REFERENCES

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.

Coltheart, M. Lexical access in simple reading task. In G. Underwood (Ed.), Strategies of information processing. London: Academic Press, 1978.

Feldman, L. Visual word recognition in Serbo-Croatian is primarily phonological. Unpublished doctoral dissertation, University of Connecticut, 1980.

Feldman, L. B. Visual word recognition in Serbo-Croatian is necessarily phonological. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, this volume.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. Representation of inflected nouns in the internal lexicon. Memory & Cognition, 1980, 8, 415-423.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. Language and Speech, 1978, 21, 142-165.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Taft, M. Lexical access via an orthographic code: The Basic Orthographic Syllable Structure (BOSS). Journal of Verbal Learning and Verbal Behavior, 1979, 18, 21-40. (a)

Taft, M. Recognition of affixed words and the word frequency effect. Memory & Cognition, 1979, 7, 263-272. (b)

FOOTNOTE

[1]Although some of the critical words in the present experiment share their dative singular form with other declined forms of other words in the language, any possible influence should be the same for both alphabets.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Baer, T. Observation of vocal fold vibration: Measurement of excised larynges. In K. N. Stevens & M. Hirano (Eds.), Vocal fold physiology. Tokyo: University of Tokyo Press, 1981, 119-133.

Bell-Berti, F., & Harris, K. S. A temporal model of speech production. Phonetica, 1981, 38, 9-21.

Best, C. T., & Harris, L. J. Childhood. In H. E. Fitzgerald & T. Carr (Eds.), Annual editions in human development 1981/82. Springfield, Ill.: Dushkin, 1981.

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 1981, 29(3), 191-211.

Fowler, C. A. Production and perception of coarticulation among stressed and unstressed vowels. Journal of Speech and Hearing Research, 1981, 46, 127-139.

Kelso, J. A. S. Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance IX. Hillsdale, N.J.: Erlbaum, in press.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. Patterns of human interlimb coordination emerge from the properties of non-linear oscillators: Theory and data. Journal of Motor Behavior, in press.

Kelso, J. A. S., & Reed, E. Motivating muscles: The problem of action. Contemporary Psychology, in press.

Kelso, J. A. S., Tuller, B., & Harris, K. S. A "dynamic pattern" perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech. Berlin: Springer-Verlag, in press.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of human movement control and coordination. London, New York: J. Wiley, in press.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. Orthography and the beginning reader. In J. F. Kavanagh & R. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, Md.: University Park Press, 1980, 137-153.

Raphael, L. J., Dorman, M. F., & Liberman, A. M. On defining the vowel duration that cues voicing in final position. Language and Speech, 1980, 23(3), 297-308.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. Speech perception without traditional speech cues. Science, 1981, 212, 947-950.

Repp, B. H. On levels of description in speech research. Journal of the Acoustical Society of America, 1981, 69, 1462-1464.

Repp, B. H. Perceptual equivalence of two kinds of ambiguous speech stimuli. Bulletin of the Psychonomic Society, in press.

Repp, B. H. Two strategies in fricative discrimination. Perception & Psychophysics, in press.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. Journal of the Acoustical Society of America, 1981, 69, 1154-1163.

Rubin, P., Baer, T., & Mermelstein, P. An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America, 1981, 70, 313-320.

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A013325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53	January - March 1978	AD A055853	ED-155-760
SR-54	April - June 1978	AD A067070	ED-161-096
SR-55/56	July - December 1978	AD AC65575	ED-166-757
SR-57	January - March 1979	AD AC83179	ED-170-823
SR-58	April - June 1979	AD A077663	ED-178-967
SR-59/60	July - December 1979	AD AC82034	ED-181-525
SR-61	January - March 1980	AD A085320	ED-185-636
SR-62	April - June 1980	AD A095062	ED-196-099
SR-63/64	July - December 1980	AD A095860	ED-197-416
SR-65	January - June 1981	AD A099958	**

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Haskins Laboratories 270 Crown Street New Haven, Connecticut 06510	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research, No. 66, April-June, 1981.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)

Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June, 1981	310	507

8a. CONTRACT OR GRANT NO. HD-01994 NS13617 N01-HD-1-2420 G-80-0178 RR-05596 MCS79-16177 PRF8006144 NS13870	9a. ORIGINATOR'S REPORT NUMBER(S) SR-66 (1981) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
N/A	See No. 8

13. ABSTRACT

This report (1 April-30 June) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

-Electromyography as a technique for laryngeal investigation
-Investigation of the phonatory mechanism
-Phonetic perception of sinusoidal signals: Effects of amplitude variation
-Memory for item order and phonetic recoding in the beginning reader
-Perceptual equivalence of two kinds of ambiguous speech stimuli
-Producing relatively unfamiliar speech gestures: A synthesis of perceptual targets and production rules
-Orthographic variations and visual information processing
-Visual word recognition in Serbo-Croatian is necessarily phonological
-Word recognition with mixed-alphabet forms
-Intra- versus inter-language Stroop effects in two types of writing systems
-Categorical perception of English /r/ and /l/ by Japanese bilinguals
-Influence of vocalic context on perception of the [ʃ]-[s] distinction: V. Two ways of avoiding it
-Grammatical priming of inflected nouns
-An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED

Security Classification

S/N 0101-807-6811 *This document contains no information not freely available to the general public. It is distributed primarily for library use.

A-31408

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception:						
categorical, /r/ and /l/, Japanese						
sinusoids, phoneme						
distinctions, /ʃ/ and /s/, context effects						
stimulus equivalence, ambiguity						
Speech Articulation:						
phonation						
production rules, perceptual targets						
larynx, electromyography						
Reading:						
Stroop effects, writing systems, cross language						
word recognition, Serbo-Croatian, phonology, alphabet						
orthography, syllables, evaluation						
orthography, information processing						
memory, phonetic recoding, order						
grammatical priming, nouns						